GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37 ; Search time 16.556 Seconds
(without alignments)
2099.101 Million cell updates/sec

Title: US-09-874-162A-5
Perfect score: 3885
Sequence: 1 MAPQKHGGGGGGGSGPSAGS.....KALETDSVSGVSKQSKKQKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1177	30.3	900	1	SU12_DROME	Q9njg9 drosophila
2	236	6.1	631	1	EMF2_ARATH	Q816y4 arabidopsis
3	217	5.6	440	1	VRN2_ARATH	Q8w5b1 arabidopsis
4	190	4.9	692	1	FIS2_ARATH	Q9znt9 arabidopsis
5	163.5	4.2	1709	1	CHD1_HUMAN	O14646 homo sapien
6	158.5	4.1	3969	1	HRX_HUMAN	Q03164 homo sapien
7	157	4.0	622	1	K1CI_HUMAN	P35527 homo sapien
8	153.5	4.0	1711	1	CHD1_MOUSE	P40201 mus musculu
9	153	3.9	549	1	DSX_DROME	P23023 drosophila
10	153	3.9	627	1	K2C1_MOUSE	P04104 mus musculu
11	151	3.9	464	1	IFE_BRALA	Q04948 branchiosto
12	146.5	3.8	1235	1	TRK1_YEAST	P12685 saccharomyc
13	146.5	3.8	1639	1	MSP1_PLAFW	P04933 plasmodium
14	146	3.8	2230	1	GOG4_HUMAN	Q13439 homo sapien
15	145	3.7	643	1	K2C1_HUMAN	P04264 homo sapien
16	145	3.7	1744	1	TANA_XENLA	Q01550 xenopus lae
17	144.5	3.7	1875	1	MLP1_YEAST	Q02455 saccharomyc
18	142.5	3.7	474	1	SOX4_HUMAN	Q06945 homo sapien
19	142.5	3.7	2038	1	FSH_DROME	P13709 drosophila
20	140	3.6	349	1	GRP_ARATH	P27483 arabidopsis
21	140	3.6	569	1	K1CJ_MOUSE	P02535 mus musculu
22	140	3.6	843	1	MNB_DROME	P49657 drosophila
23	139.5	3.6	458	1	K1CM_HUMAN	P13646 homo sapien
24	139.5	3.6	912	1	UBP3_YEAST	Q01477 saccharomyc
25	139.5	3.6	1085	1	YAFA_SCHPO	Q09863 schizosacch
26	139	3.6	252	1	GRP1_PHAVU	P10495 phaseolus v
27	139	3.6	419	1	K1C4_XENLA	P05781 xenopus lae
28	139	3.6	481	1	LORI_MOUSE	P18165 mus musculu
29	139	3.6	1453	1	Y373_BOVIN	Q9tu23 bos taurus
30	139	3.6	1833	1	ZEP2_HUMAN	P31629 homo sapien
31	138	3.6	1701	1	MSP1_PLAFM	P08569 plasmodium
32	137	3.5	641	1	EBN1_EBV	P03211 epstein-bar
33	136	3.5	384	1	GRP1_PETHY	P09789 petunia hyb
34	136	3.5	1701	1	MSP1_PLAFF	P13819 plasmodium
35	135.5	3.5	183	1	GRP2_ORYSA	P29834 oryza sativ
36	135.5	3.5	1142	1	GIN4_YEAST	Q12263 saccharomyc
37	135	3.5	1953	1	BNI1_YEAST	P41832 saccharomyc
38	135	3.5	3038	1	TRIO_HUMAN	O75962 homo sapien
39	134.5	3.5	440	1	FXGA_CHICK	Q98937 gallus gall
40	134.5	3.5	629	1	K2C3_HUMAN	P12035 homo sapien
41	134.5	3.5	971	1	Y228_BORBU	O51246 borrelia bu
42	134.5	3.5	1790	1	USO1_YEAST	P25386 saccharomyc
43	134	3.4	465	1	GRP2_PHAVU	P10496 phaseolus v
44	134	3.4	593	1	K1CJ_HUMAN	P13645 homo sapien
45	133	3.4	526	1	K1CJ_BOVIN	P06394 bos taurus

ALIGNMENTS

RESULT 1
SU12_DROME
ID SU12_DROME STANDARD; PRT; 900 AA.
AC Q9NJG9; Q8T9D8; Q9VW55;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycomb protein Su(z)12 (Suppressor 12 of zeste protein).
GN SU(Z)12 OR CG8013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT
RP SU(Z)12-2.
RX MEDLINE=21430867; PubMed=11546753;
RA Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,
RA Rasmuson-Lestander A., Mueller J.;
RT "Su(z)12, a novel Drosophila Polycomb group gene that is conserved in
RT vertebrates and plants.";
RL Development 128:3371-3379(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RC   STRAIN=Berkeley;
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [5]
RP   IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAF1 AND ESC,
RP   AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX   MEDLINE=22296673; PubMed=12408863;
RA   Czermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V.;
RT   "Drosophila enhancer of Zeste/ESC complexes have a histone H3
RT   methyltransferase activity that marks chromosomal Polycomb sites.";
RL   Cell 111:185-196(2002).
RN   [6]
RP   IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAF1 AND ESC, AND
RP   METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX   MEDLINE=22296674; PubMed=12408864;
RA   Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,
RA   Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;
RT   "Histone methyltransferase activity of a Drosophila Polycomb group
RT   repressor complex.";
RL   Cell 111:197-208(2002).
CC   -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
CC       forming multiprotein complexes, which are required to maintain the
CC       transcriptionally repressive state of homeotic genes throughout
CC       development. PcG proteins are not required to initiate repression,
CC       but to maintain it during later stages of development. They
CC       probably act via the methylation of histones, rendering chromatin
CC       heritably changed in its expressibility. Component of the Esc/E(z)
CC       complex, which methylates Lys-9 and Lys-27 residues of histone H3.
CC       Despite the presence of a zinc-finger, it does not bind directly
CC       to DNA, the Esc/E(z) complex being probably recruited to DNA by
CC       Pho. The Esc/E(z) complex is necessary but not sufficient to
CC       recruit a functional PcG repressive complex that represses target
CC       genes, suggesting that the recruitment of the distinct PRC1
CC       complex is also required to allow a subsequent repression.
CC   -!- SUBUNIT: Component of the Esc/E(z) complex, composed of Esc, E(z),
CC       Su(z)12, Rpd3, Caf1 and probably Pho. This complex is distinct
CC       from the PRC1 complex, which contains many other PcG proteins like
CC       Pc, Ph, Psc, Su(z)2. The two complexes however cooperate and
CC       interact together during the first 3 hours of development to
CC       establish PcG silencing.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1; Synonyms=B;
CC         IsoId=Q9NJG9-1; Sequence=Displayed;
CC       Name=2; Synonyms=A;
CC         IsoId=Q9NJG9-2; Sequence=VSP_007033, VSP_007034;
CC         Note=No experiment confirmation available;
CC   -!- SIMILARITY: Belongs to the VEFS (VRN2-EMF2-FIS2-SU(Z)12) family.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF149047; AAF73149.1; -.
DR   EMBL; AE003515; AAF49094.2; -.
DR   EMBL; AE003515; AAN11641.1; -.
DR   EMBL; AY069809; AAL39954.1; -.
DR   FlyBase; FBgn0020887; Su(z)12.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW   Transcription regulation; Repressor; Developmental protein;
KW   Nuclear protein; Metal-binding; Zinc; Zinc-finger;
KW   Alternative splicing.
FT   ZN_FING      411    434       C2H2-TYPE.
FT   DOMAIN       527    603       VEFS-BOX.
FT   DOMAIN       355    366       ASN-RICH.
FT   DOMAIN       699    879       SER-RICH.
FT   VARSPLIC     806    855       NTVLNKRQRYSDGSPGTGIGNGHGGGSGSGANRNKSNNHSL
FT                                 PATSNNASS -> VEQAADAPEVLTHSDNAVDVGIIDDECG
FT                                 GFGAVGVMNGVASPVANNCVGN (in isoform 2).
FT                                 /FTId=VSP_007033.
FT   VARSPLIC     856    900       Missing (in isoform 2).
FT                                 /FTId=VSP_007034.
FT   MUTAGEN      274    274       G->D: IN SU(Z)12-2; INDUCES LARVAL
FT                                 LETHALITY WHEN HOMOZYGOUS.
SQ   SEQUENCE   900 AA;  100104 MW;  53BA0D83C49EC92F CRC64;

  Query Match             30.3%;  Score 1177;  DB 1;  Length 900;
  Best Local Similarity   34.8%;  Pred. No. 9.1e-61;
  Matches  281;  Conservative  136;  Mismatches  225;  Indels  166;  Gaps   26;

Qy         43 GGSCGG--GGSYSASSSSSAAAAAGAAVLPVK--KPKMEHVQADHELFLQAFEKPTQIYR 98
              |  :  |  | :: |  :|:    ||: | |    : |:   | : ||||||||||||||
Db         15 GSAANGIIGLTHGAPDASN----AGSTVPPTAEGQVKLNGHQQEQELFLQAFEKPTQIYR 70

Qy         99 FLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSA-HL 157
              :|| |:   ||||:|||:||  | || | || :|:|: ||  :      |:| ::|  :|
Db         71 YLRNRHETNPIFLNRTLSYMKERMSRNNKKRISFQVNSMLESI-----TQKSEAVSQNYL 125

Qy        158 QLTFTGFFHKNDKPSPNSENEQ----------NSVTLEVLLVKVCHKKRKDVSCPIRQV 206
               : :     |       |   |              ||::|  | |:   |||| :   :::
Db        126 HVIYDSLHEKLPARLDNESGEDLLQEQLLCEAGESVSVETTLYKITRSKRKDSTLDFQEL 185

Qy        207 PTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEP-SNSHMVKSYSLLFRVTRPGRREFNGM 265
               :   |:   |   : : |   ::::      |    | :  | ||||:          :
Db        186 LSKCSQIVYNP---KDRVGEHATISIPLQTMRPMGEQHTL--YKLLFRIK---------V 231

Qy        266 INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEEC 325
               ::  |    | | | |   ||:| : || | ::: :::|:   : :||||  :| :
Db        232 LSPSTCN--DENAETPP-NKRSRPN-EKMFGSELILYEKSSGF-ITEGEYEAMLQPLNST 286

Qy        326 PI---SKKRATWETILDGKRLP---PFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLAT 379
               |   | |: ||||: |   :|    :: : | | |:| |     :|::    |: |
Db        287 SIKSFSPKKCTWETMPD-SYIPLSLTYDVYQQSPMLKFHLTL---SNEQLPEMISAPELQ 342

Qy        380 R-------------NSESLHQEN-----KPGSVKPTQTIAVKESLTTDLQTRKEKDTPNE 421
               |             |: : :  |     | ||     |:                | ||
Db        343 RYVQHLDAVAEMNYNNNNYNNNNNCSGLKNGSGGGNSTVC-------------KTTP-- 386

Qy        422 NRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYH 481
```

```
                 : ::| | |:|:|||||||||   :|:|||| |:| :||:||||||||||:|| | |
Db        387 --EHIQIVYNFMYSNNTRQQTEYTQELNCPWCGLDCLRLYALLKHLKLCHARFNFTYQPA 444

Qy        482 PKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRN-GPVKRTPITHILVCRPKRTKASM 540
                |||||:||: |||||||:| |:    | :|:|  |||:|| :| ::|||||:| |  :
Db        445 GSGARIDVTINDAYDGSYAGSPYDLAGPSGSSFARTCGPVRRTSVTSLMVCRPRRQKTCL 504

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
                |||| :: |:  ||:| :|||||| |::|||| |:|:::||| | || |||:||| |
Db        505 DEFLELDEDEISNQRSYITGHNRLYHHTETCLPVHPKELDIDSEGESDPLWLRQKTIQMI 564

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
               :|||||||||||:|||||||||:|||: | |:  || :|::  | :|::||| |||:||:
Db        565 DEFSDVNEGEKELMKLWNLHVMRHGFVGDCQLPIACEMFLDAKGTEIVRKNLYRNFILHM 624

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGES-------------------------- 693
                |: |: ||:  :: | | ||: :  |   |:
Db        625 CSLFDYGLIAAETVYKTVQKLQGLLSKYAAGQELMQRQREEQLKYWLDVGMHKKQEDPKT 684

Qy        694 -------ASPANEEITE---------------------------------EQNGTANG 711
                    | ||::  |                                    :  || ||
Db        685 LKSPQKPAPPADQASTSSASTSGSGSGSSSMQPPKRMPAHLKRGSAASSPGVQSKGTENG 744

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
                :  ||      : :   |:|:|  | |
Db        745 TNGSNS-----SSSNSKNVAKKSADQPL 767


RESULT 2
EMF2_ARATH
ID   EMF2_ARATH     STANDARD;      PRT;   631 AA.
AC   Q8L6Y4; Q93V59; Q9LU50; Q9LU51;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polycomb protein EMBRYONIC FLOWER 2.
GN   EMF2 OR AT5G51230/AT5G51240 OR MWD22.18/MWD22.19.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=cv. Columbia;
RX   PubMed=11701882;
RA   Yoshida N., Yanai Y., Chen L., Kato Y., Hiratsuka J., Miwa T.,
RA   Sung Z.R., Takahashi S.;
RT   "EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates
RT   shoot development and flowering in Arabidopsis.";
RL   Plant Cell 13:2471-2481(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   SSP consortium (Salk/Stanford/PGEC).";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Polycomb group (PcG) protein. Involved in flowering
CC       processes by repressing unknown target genes and preventing
CC       reproductive development. PcG proteins act by forming multiprotein
CC       complexes, which are required to maintain the transcriptionally
CC       repressive state of homeotic genes throughout development. PcG
CC       proteins are not required to initiate repression, but to maintain
CC       it during later stages of development. They probably act via the
CC       methylation of histones, rendering chromatin heritably changed in
CC       its expressibility.
CC   -!- SUBUNIT: In plants, PcG complexes are probably composed of a
CC       member of the EZ family (CLF or MEA), FIE, and a member of the
CC       VEFS family (FIS2, VRN2 or EMF2) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q8L6Y4-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q8L6Y4-2; Sequence=VSP_007456;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Widely expressed throughout the life cycle
CC       with higher levels in proliferating tissues. Expressed in both
CC       vegetative and the reproductive shoot meristems.
CC   -!- DEVELOPMENTAL STAGE: Expressed in the developing embryos and
CC       endosperm, then decreases when embryos mature and soon after
CC       cellularization in the endosperm. After germination, it is
CC       expressed in the shoot apical meristems (SAMs), leaf primordia,
CC       and young leaves. In the reproductive shoots, it is expressed in
CC       both the influorescence and floral meristems. Later, it is
CC       expressed in floral organ primordia. In coflorescences, it is
CC       expressed in SAMs and lateral organs. In roots, it is expressed in
CC       root tips.
CC   -!- SIMILARITY: Belongs to the VEFS (VRN2-EMF2-FIS2-SU(Z)12) family.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB053171; BAB58956.1; -.
DR   EMBL; AB053262; BAB58957.1; -.
DR   EMBL; AB023044; BAA97386.1; ALT_SEQ.
DR   EMBL; AB023044; BAA97387.1; ALT_SEQ.
DR   EMBL; AY140086; AAM98227.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; Repressor; Flowering; Nuclear protein;
KW   Zinc; Zinc-finger; Metal-binding; Alternative splicing.
FT   ZN_FING      324    347       C2H2-TYPE.
FT   DOMAIN       505    583       VEFS-BOX.
FT   DOMAIN        32     35       POLY-GLU.
FT   DOMAIN       289    292       POLY-SER.
FT   DOMAIN       396    399       POLY-ARG.
FT   VARSPLIC     392    396       Missing (in isoform 2).
FT                                 /FTId=VSP_007456.
SQ   SEQUENCE   631 AA;  71680 MW;  F36FE92D0F62E610 CRC64;

  Query Match             6.1%;  Score 236;  DB 1;  Length 631;
  Best Local Similarity   19.4%;  Pred. No. 1.4e-06;
  Matches  137;  Conservative  104;  Mismatches  258;  Indels  208;  Gaps   24;

Qy         77 EHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDD 136
                | : |:    | |: || ::|  :: | :  |:|| | | |     :  |:   |
Db         34 EEIAAEES--LAAYCKPVELYNIIQRRAIRNPLFLQRCLHYKIEAKHKRRIQMTVFL--- 88

Qy        137 MLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
                          | :: | :|                    |    | :|| ::   |
Db         89 ------------SGAIDAGVQ------------------TQKLFPLYILLARLVSPK- 115

Qy        197 KDVSCPIRQ-----------VPTGKKQVP---------LIPDLN----QTKPGN----FP 228
                  |: :           : ||   |          |:||:|    : | |:    |
```

```
Db        116 -----PVAEYSAVYRFSRACILTGGLGVDGVSQAQANFLLPDMNRLALEAKSGSLAILFI 170

Qy        229 SLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNR 288

Db        171 SFAGAQNSQFGIDSGKIHS---------------GNIGGHC-----LWSKIPLQ----- 204

Qy        289 EDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFET 348

Db        205 --------SLYASWQKSPNMDLGQRVDTVSLVEMQPCFIKLKSMSEEKCVS--------- 247

Qy        349 FSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTT 408

Db        248 -IQVPSNPLT----------SSSPQQVQVTISAEEVGSTEKSPYSSFSYNDIS-SSSLLQ 295

Qy        409 DLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLK 468

Db        296 IIRLR----TGN-------VVFNYRYYNNKLQKTEVTEDFSCPFCLVKCASFKGLRYHLP 344

Qy        469 LCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIH-RQPGFAF------------- 514

Db        345 STHDLLNFEFWVTEEFQAVNVSLKTETMISKV-NEDDVDPKQQTFFFSSKKFRRRRQKSQ 403

Qy        515 ---SRNGP--------VKRTPITHILVCRPKR---------------------------- 535

Db        404 VRSSRQGPHLGLGCEVLDKTDDAHSVRSEKSRIPPGKHYERIGGAESGQRVPPGTSPADV 463

Qy        536 ------------TKASMSEFLESEDGEVEQQ--RTYSSGHNRLYFHSDTCLPLRPQEM-- 579

Db        464 QSCGDPDYVQSIAGSTMLQFAKTRKISIERSDLRNRSLLQKRQFFHSHRAQPMALEQVLS 523

Qy        580 EVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLF 639

Db        524 DRDSEDEVDDDVADFEDRRMLDDFVDVTKDEKQMMHMWNSFVRKQRVLADGHIPWACEAF 583

Qy        640 VENYGQKIIK-KNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ 685

Db        584 SRLHGPIMVRTPHLIWCWRVFMVKLWNHGLLDARTMNNCNTFLEQLQ 630
```

```
RESULT 3
VRN2_ARATH
ID   VRN2_ARATH     STANDARD;      PRT;   440 AA.
AC   Q8W5B1; O23524; O23525; Q8W5B2; Q94CF5;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polycomb group protein VERNALIZATION 2.
GN   VRN2 OR AT4G16845 OR FCAALL.23 OR DL4450W.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), FUNCTION, AND CHARACTERIZATION.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=21575875; PubMed=11719192;
RA   Gendall A.R., Levy Y.Y., Wilson A., Dean C.;
RT   "The VERNALIZATION2 (VRN2) gene mediates the epigenetic regulation of
RT   vernalization in Arabidopsis.";
RL   Cell 107:525-535(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=98121113; PubMed=9461215;
RA   Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,
RA   Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA   Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,
RA   Wedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T., Terryn N.,
RA   Gielen J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,
RA   Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,
RA   Kotter P., Entian K.-D., Rieger M., Schaefer M., Funk B.,
RA   Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,
RA   Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P.,
RA   Piravandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,
RA   Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansorge W.,
RA   Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,
RA   Klosterman S., Schueller C., Chalwatzis N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   Nature 391:485-488(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   SSP consortium (Salk/Stanford/PGEC).";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Polycomb group (PcG) protein. Plays a central role in
CC       vernalization by maintaining repressed the homeotic gene FLC, a
CC       floral repressor, after a cold treatment. PcG proteins act by
CC       forming multiprotein complexes, which are required to maintain the
CC       transcriptionally repressive state of homeotic genes throughout
CC       development. PcG proteins are not required to initiate repression,
CC       but to maintain it during later stages of development. They
CC       probably act via the methylation of histones, rendering chromatin
```

```
CC       heritably changed in its expressibility.
CC   -!- SUBUNIT: Probable component of a PcG complex. In plants, PcG
CC       complexes are probably composed of a member of the EZ family (CLF
CC       or MEA), FIE, and a member of the VEFS family (FIS2, VRN2 or EMF2)
CC       (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q8W5B1-1; Sequence=Displayed;
CC       Name=2; Synonyms=VRN2';
CC         IsoId=Q8W5B1-2; Sequence=VSP_007457, VSP_007458;
CC   -!- TISSUE SPECIFICITY: Weakly expressed. Expressed both during, and
CC       in the absence of vernalization.
CC   -!- SIMILARITY: Belongs to the VEFS (VRN2-EMF2-FIS2-SU(Z)12) family.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
CC       erroneous gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF284500; AAL32135.1; -.
DR   EMBL; AF284501; AAL32136.1; -.
DR   EMBL; Z97342; CAB10457.1; ALT_SEQ.
DR   EMBL; AL161545; CAB80955.1; ALT_SEQ.
DR   EMBL; AY034902; AAK59409.1; -.
DR   EMBL; AY063047; AAL34221.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; Repressor; Nuclear protein; Zinc;
KW   Zinc-finger; Metal-binding; Alternative splicing.
FT   ZN_FING      86    111       C2H2-TYPE.
FT   DOMAIN      156    163       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      267    345       VEFS-BOX.
FT   DOMAIN      398    402       POLY-ASN.
FT   DOMAIN      411    416       POLY-ASN.
FT   VARSPLIC    102    107       GLQFHL -> VGNYYN (in isoform 2).
FT                                /FTId=VSP_007457.
FT   VARSPLIC    108    440       Missing (in isoform 2).
FT                                /FTId=VSP_007458.
FT   CONFLICT     81     81       R -> K (IN REF. 1).
FT   CONFLICT    121    121       L -> S (IN REF. 1).
FT   CONFLICT    185    185       A -> T (IN REF. 1).
FT   CONFLICT    193    193       A -> T (IN REF. 1).
FT   CONFLICT    239    239       T -> S (IN REF. 1).
FT   CONFLICT    344    344       V -> A (IN REF. 1).
FT   CONFLICT    355    355       C -> R (IN REF. 1).
FT   CONFLICT    394    398       TSVTN -> SSDTTTT (IN REF. 1).
FT   CONFLICT    406    406       H -> R (IN REF. 1).
FT   CONFLICT    425    425       K -> N (IN REF. 1).
FT   CONFLICT    440    440       K -> KVIK (IN REF. 1).
SQ   SEQUENCE   440 AA;  50623 MW;  0E0AB2C55517BE6F CRC64;

  Query Match             5.6%;  Score 217;  DB 1;  Length 440;
  Best Local Similarity   21.2%;  Pred. No. 1.1e-05;
  Matches   80;  Conservative   62;  Mismatches  140;  Indels   96;  Gaps   11;

Qy        413 RKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHS 472
              || |     :|    : : :   ||| |:|| |:|  ||:|:: |     |  ||   |
Db         58 RKRR-----SRSTGMVVFNYKDCNNTLQRTEVREDCSCPFCSMLCGSFKGLQFHLNSSHD 112

Qy        473 RFIF-----------------NYVYHPKGARIDV----------------------SI 491
               | |                 ::::  :|:  |                      :
Db        113 LFEFEFKLLEEYQTVNVSVKLNSFIFEEEGSDDDKFEPFSLCSKPRKRRQRGGRNNTRRL 172

Qy        492 NECY-------------------DGSYA-GNPQDIHRQPGFAFSRNGPVKRTPITH--- 527
```

```
               |:                    ||:   | |:       |  : | |     | |
Db        173 KVCFLPLDSPSLANGTENGIALLNDGNRGLGYPEATELAGQFEMTSNIP---PAIAHSSL 229

Qy        528 -----ILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEM--E 580
                  :::       |:.:  | :|  |    |::     | ::||    |:  :::  :
Db        230 DAGAKVILTTEAVVPATKTRKLSAERSEA---RSHLLLQKRQFYHSHRVQPMALEQVMSD 286

Qy        581 VDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFV 640
              ||||| | :    :    :::| |||: ||: | |||  | |   |||  :: || :|
Db        287 RDSEDEVDDDVADFEDRQMLDDFVDVNKDEKQFMHLWNSFVRKQRVIADGHISWACEVFS 346

Qy        641 ENYGQKI-IKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANE 699
              | :::    :|   : | |: : :  |:   :|:   | |   :             |
Db        347 RFYEKELHCYSSLFWCWRLFLIKLWNHGLVDSATINNCNTILENCR-----------NT 394

Qy        700 EITEEQNGTANGFSEINS 717
              :|   | : :  |: |:
Db        395 SVTNNNNNSVDHPSDSNT 412


RESULT 4
FIS2_ARATH
ID   FIS2_ARATH     STANDARD;      PRT;   692 AA.
AC   Q9ZNT9; Q9ZQP0;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polycomb group protein FERTILIZATION-INDEPENDENT SEED 2.
GN   FIS2 OR AT2G35670 OR T20F21.14.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A., MUTANT FIS2-4, AND VARIANTS.
RC   STRAIN=cv. Landsberg erecta; TISSUE=Siliques;
RX   MEDLINE=99093530; PubMed=9874812;
RA   Luo M., Bilodeau P., Koltunow A., Dennis E.S., Peacock W.J.,
RA   Chaudhury A.;
RT   "Genes controlling fertilization-independent seed development in
RT   Arabidopsis thaliana.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:296-301(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [3]
RP   SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX   PubMed=10962025;
RA   Luo M., Bilodeau P., Dennis E.S., Peacock W.J., Chaudhury A.;
RT   "Expression and parent-of-origin effects for FIS2, MEA, and FIE in the
RT   endosperm and embryo of developing Arabidopsis seeds.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:10637-10642(2000).
CC   -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
CC       forming multiprotein complexes, which are required to maintain the
CC       transcriptionally repressive state of homeotic genes throughout
CC       development. PcG proteins are not required to initiate repression,
CC       but to maintain it during later stages of development. They
CC       probably act via the methylation of histones, rendering chromatin
CC       heritably changed in its expressibility. Required to prevent the
CC       proliferation of the central cell by repressing unknown target
```

```
CC       genes before fertilization.
CC   -!- SUBUNIT: Probably indirectly associated with FIE and/or MEA. In
CC       plants, PcG complexes are probably composed of a member of the EZ
CC       family (CLF or MEA), FIE, and a member of the VEFS family (FIS2,
CC       VRN2 or EMF2) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Weakly expressed. Expressed in late siliques.
CC   -!- DEVELOPMENTAL STAGE: Expressed maternally and zygotically.
CC       Expressed in the central cell before fertilization, and in the
CC       endosperm after fertilization, then decreases before the time of
CC       endosperm cellularization but continues in the chalazal cyst.
CC   -!- POLYMORPHISM: In cv. Columbia, the sequence differs from that
CC       shown due to a deletion in the genomic sequence that covers the
CC       Leu-443 to Ile-502 region.
CC   -!- SIMILARITY: Belongs to the VEFS (VRN2-EMF2-FIS2-SU(Z)12) family.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF096095; AAD09104.1; -.
DR   EMBL; AF096096; AAD09105.1; -.
DR   EMBL; AC006068; AAD15448.2; ALT_SEQ.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; Repressor; Nuclear protein;
KW   Developmental protein; Repeat; Zinc-finger; Metal-binding;
KW   Polymorphism.
FT   ZN_FING      11     34       C2H2-TYPE.
FT   DOMAIN      122    502       12 X APPROXIMATE REPEAT A.
FT   REPEAT      122    143       A-1.
FT   REPEAT      161    183       A-2.
FT   REPEAT      184    206       A-3.
FT   REPEAT      207    228       A-4.
FT   REPEAT      229    250       A-5.
FT   REPEAT      299    320       A-6.
FT   REPEAT      338    360       A-7.
FT   REPEAT      361    381       A-8.
FT   REPEAT      399    421       A-9.
FT   REPEAT      422    442       A-10.
FT   REPEAT      459    480       A-11.
FT   REPEAT      481    502       A-12.
FT   DOMAIN      144    519       7 X APPROXIMATE REPEAT B.
FT   REPEAT      144    160       B-1.
FT   REPEAT      251    267       B-2.
FT   REPEAT      282    298       B-3.
FT   REPEAT      321    337       B-4.
FT   REPEAT      382    398       B-5.
FT   REPEAT      443    458       B-6.
FT   REPEAT      503    519       B-7.
FT   DOMAIN      543    621       VEFS-BOX.
FT   VARIANT     269    269       G -> V (in cv. Columbia).
FT   VARIANT     442    501       Missing (in cv. Columbia).
FT   VARIANT     554    554       M -> I (in cv. Columbia).
FT   MUTAGEN     637    692       MISSING: IN FIS2-4; INDUCES ATROPHY OF
FT                                PARTIALLY DEVELOPING SEEDS BEYOND THE
FT                                GLOBULAR STAGE.
SQ   SEQUENCE   692 AA;  77666 MW;  2A3F61B961676497 CRC64;

  Query Match             4.9%;  Score 190;  DB 1;  Length 692;
  Best Local Similarity   18.8%;  Pred. No. 0.00071;
  Matches  126;  Conservative   99;  Mismatches  267;  Indels  178;  Gaps   24;

Qy        158 QLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKV------------CHKKRKDVSCPIR 204
              ||:   |  ||       ::  |   | |||:::            |    :|| | |
Db         74 QLSPLTFCSKNRNQRRQRDDSNNVKKLNVLLMELDLDDLPRGTENDSTHVNDDNVSSPPR 133

Qy        205 -------------QVPTGKKQVPLIPDLNQTKPGNF--PSLAVSSNE----------- 236
                          |:   :   | :| :|    ||   |  | || |
Db        134 AHSSEKISDILTTTQLAIAESSEPKVPHVND---GNVSSPPRAHSSAEKNESTHVNDDDD 190

Qy        237 -FEPSNSHMVKSYSLLF----RVTRPGRREFNGMINGETNENIDVNEE---LPARRKRNR 288
               |  :| ::          :: | :     : : ||:  :|:|    | | : ::
Db        191 VSSPPRAHSLEKNESTHVNEDNISSPPKAH-----SSKKNESTHMNDEDVSFPPRTRSSK 245

Qy        289 EDGEKTFVAQMTVFDK--------NRRLQLLDGEYEVAMQEMEECPISKKRATWETILDG 340
              |  :     |  : :          :|| ||    |              | | |   | :
Db        246 ETSDILTTTQPAIVEPSEPKVRRGSRRKQLYAKRY------------KARETQPAIAES 292

Qy        341 KRLPPF----ETFSQGP-------------TLQFTLRWTGE-----TNDKSTAPIAKPLA 378
                  |  |  |              | |  :  : |      ||:: :   :  :
Db        293 SEPKVLHVNDENVSSPPEAHSLEKASDILTTTQPAIAESSEPKVPHVNDENVSSTPRAHS 352

Qy        379 TRNSESLHQ--ENKPGSVKPTQTIAVKESLTTDLQTRKEKDTP------NENRQKLRIFY 430
              :: ::|  :  :| | |   |   :    : |||   | |  |      ::|     :
Db        353 SKKNKSTRKNVDNVPSPPKTRSSKKTSDILTTTQPTIAESSEPKVRHVNDDNVSSTPRAH 412

Qy        431 QFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVS 490
               | :||:     |::  |  | : :|  ::|   :              |  |  :
Db        413 SSKKNKSTRKND---DNIPSPPKTRSSKKTSNILTRTQ------------PAIAESEPK 456

Qy        491 INECYDGSYAGNP--------QDIHRQPGFA--------------FSRNGPVKRTPITH 527
              :   |  : |         :  |::  |                :  | |  |
Db        457 VPHVNDDKVSSTPRAHSSKKNKSTHKKDDNASLPPKTRSSKKTSDILATTQPAKAEP--- 513

Qy        528 ILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEM--EVDSED 585
               || |: |  : | :|  | ::          | ::|| |  |:  :::    |||:
Db        514 ---SEPKVTRVSRRKELHAERCEAKRLERLK---GRQFYHSQTMQPMTFEQVMSNEDSEN 567

Qy        586 EKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVE---- 641
              | |   |     ::|    |:: ||  | |||: | |   |||  :  ||  | :
Db        568 ETDDYALDISERLRLERLVGVSKEEKRYMYLWNIFVRKQRVIADGHVPWACEEFAKLHKE 627

Qy        642 --------NYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGES 693
                    ::  :: :  |  | ::   : |    | : : |:|         :   | :
Db        628 EMKNSSSFDWWWRMFRIKLWNNGLICAKTFHKCTTILLSNSDEA--------GQFTSGSA 679

Qy        694 ASPANEEITE 703
              |:  |::  |
Db        680 ANANNQQSME 689


RESULT 5
CHD1_HUMAN
ID   CHD1_HUMAN     STANDARD;      PRT;  1709 AA.
AC   O14646;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
GN   CHD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97470991; PubMed=9326634;
RA   Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT   "Characterization of the CHD family of proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC       IMPORTANT ROLE IN GENE REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: Contains 2 chromo domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF006513; AAB87381.1; -.
DR   HSSP; P23197; 1AP0.
DR   Genew; HGNC:1915; CHD1.
DR   MIM; 602118; -.
DR   GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.
DR   GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00385; chromo; 2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN        1     70       SER-RICH.
FT   DOMAIN      117    137       SER-RICH.
FT   DOMAIN      272    364       CHROMO 1.
FT   DOMAIN      389    452       CHROMO 2.
FT   NP_BIND     506    513       ATP (POTENTIAL).
FT   SITE        614    617       DEAH BOX.
FT   DOMAIN     1628   1644       3 X 5 AA REPEATS OF H-S-D-H-R.
FT   REPEAT     1628   1632       1.
FT   REPEAT     1634   1638       2.
FT   REPEAT     1640   1644       3.
SQ   SEQUENCE   1709 AA;  196517 MW;  416409C913D6A935 CRC64;

  Query Match             4.2%;  Score 163.5;  DB 1;  Length 1709;
  Best Local Similarity   20.4%;  Pred. No. 0.074;
  Matches  128;  Conservative   79;  Mismatches  174;  Indels  245;  Gaps   32;

Qy         14 SGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAVLPVKK 73
              || ::||| |   ||::  ::: |     | |    || | | | | :  :    :  |
Db         24 SGSASGSGSGSSSGSSSDGSSSQS-----GSSDSDSGSESGSQSESESDTSRENKVQAKP 78

Qy         74 PKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFK 133
              ||::  :                  | ::   |  :              |: | :|
Db         79 PKVDGAE-----------------FWKSSPSILAV------------QRSAILKK--- 104

Qy        134 VDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCH 193
                     :: : :|: |  |::           ::: | :||:  :| :      :|
Db        105 --------QQQQQQQQQHQASSN---------SGSEEDSSSSEDSDDSSS------EVKR 141

Qy        194 KKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNE---------------F 237
              || ||                |   :  |: || : | :|                 :
Db        142 KKHKD-----------------EDWQMSGSGS-PSQSGSDSESEEEREKSSCDETESDY 182

Qy        238 EPSNSHMVKSYSLLFRVTRPGRREFNG-MINGETNENIDVNEELPARRKRNREDGEKTFV 296
              || |   |||     |  :   :  ||  | |:    || :||          :| |:
Db        183 EPKNK--VKS-----RKPQNRSKSKNGKKILGQKKRQIDSSEE---------DDDEED-- 224

Qy        297 AQMTVFDKNRRLQLLDGEYEVAMQEMEE------------------------------ 324
                  :| ::|         |: :| ||
Db        225 -----YDNDKRSSRRQATVNVSYKEDEEMKTDSDDLLEVCGEDVPQPEEEEFETIERFMD 279

Qy        325 CPISKKRATWETI------LDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLA 378
              | | :| ||  |        ||    ||   :   :|: ::| | ::  :|
Db        280 CRIGRKGATGATTTIYAVEADGDPNAGFEKNKEPGEIQYLIKWKGWSHIHNT-------- 331

Qy        379 TRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN--ENRQKLRIFYQFLYNN 436
                 |:| |:|  |  |             |   :|:::|    :|     : |   ||
Db        332 WETEETLKQQNVRGMKK-----------LDNYKKKDQETKRWLKNASPEDVEY---YN- 375

Qy        437 NTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHS--RFIFNYV-YHPK---------- 483
                 | |  ||||         | | ::  : : ||  :    |  |: |
Db        376 ---CQQELTDDLH---------KQYQIVGRI-IAHSNQKSAAGYPDYYCKWQGLPYSECS 422

Qy        484 ---GARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPI--THILVCRPKRTKA 538
                 || |      | |  |                  ||||   | ||     :|  ||:
Db        423 WEDGALISKKFQACID-EY-------------FSRN-QSKTTPFKDCKVLKQRPR---- 462

Qy        539 SMSEFLESEDGEVEQQRTYSSGHNRL 564
                  |:      :::| :|  ||  |
Db        463 ----FV-----ALKKQPSYIGGHEGL 479


RESULT 6
HRX_HUMAN
ID   HRX_HUMAN      STANDARD;      PRT;  3969 AA.
AC   Q03164; Q13743; Q13744; Q14845; Q16364; Q9UMA3;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zinc finger protein HRX (ALL-1) (Trithorax-like protein).
GN   MLL OR HRX OR ALL1 OR TRX1 OR HTRX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93046667; PubMed=1423624;
RA   Tkachuk D.C., Kohler S., Cleary M.L.;
RT   "Involvement of a homolog of Drosophila trithorax by 11q23
RT   chromosomal translocations in acute leukemias.";
RL   Cell 71:691-700(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96290553; PubMed=8703835;
RA   Nilson I., Loechner K., Siegler G., Greil J., Beck J.D., Fey G.H.,
RA   Marschalek R.;
RT   "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT   translocations to chromosomal region 11q23 and acute leukaemias.";
RL   Br. J. Haematol. 93:966-972(1996).
RN   [3]
RP   SEQUENCE OF 1-1909 FROM N.A.
RX   MEDLINE=93390935; PubMed=8378076;
RA   Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
RA   Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT   "Two distinct portions of LTG19/ENL at 19p13 are involved in t(11;19)
RT   leukemia.";
RL   Oncogene 8:2617-2625(1993).
RN   [4]
RP   SEQUENCE OF 1317-2328 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93265134; PubMed=1303259;
RA   Djabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT   "A trithorax-like gene is interrupted by chromosome 11q23
RT   translocations in acute leukaemias.";
RL   Nat. Genet. 2:113-118(1992).
RN   [5]
RP   SEQUENCE OF 1251-1538 FROM N.A.
RX   MEDLINE=94215165; PubMed=8162575;
RA   Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaani O.,
RA   Saito H., Croce C.M., Canaani E.;
RT   "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT   involved in acute leukemia.";
RL   Cancer Res. 54:2326-2330(1994).
RN   [6]
RP   SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX   MEDLINE=95322025; PubMed=7598802;
RA   Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
RA   Rowley J.D., Diaz M.O.;
RT   "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT   trx zinc-finger domain, and alternative splicing.";
```

```
RL   DNA Cell Biol. 14:475-483(1995).
RN   [7]
RP   SEQUENCE OF 1212-1603 FROM N.A.
RX   MEDLINE=95315013; PubMed=7794749;
RA   Marschalek R., Greil J., Lochner K., Nilson I., Siegler G.,
RA   Zweckbronner I., Beck J.D., Fey G.H.;
RT   "Molecular analysis of the chromosomal breakpoint and fusion
RT   transcripts in the acute lymphoblastic SEM cell line with chromosomal
RT   translocation t(4;11).";
RL   Br. J. Haematol. 90:308-320(1995).
RN   [8]
RP   SEQUENCE OF 1421-1540 FROM N.A.
RX   MEDLINE=94020842; PubMed=8414518;
RA   Forster A., Rabbitts T.H.;
RT   "A method for identifying genes within yeast artificial chromosomes:
RT   application to isolation of MLL fusion cDNAs from acute leukaemia
RT   translocations.";
RL   Oncogene 8:3157-3160(1993).
RN   [9]
RP   CHROMOSOMAL TRANSLOCATION WITH GAS7.
RX   MEDLINE=20183971; PubMed=10706619;
RA   Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
RA   Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
RA   Williams T.M., Lange B.J., Felix C.A.;
RT   "Detection of leukemia-associated MLL-GAS7 translocation early during
RT   chemotherapy with DNA topoisomerase II inhibitors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
RN   [10]
RP   CHROMOSOMAL TRANSLOCATION WITH AF3P21.
RX   MEDLINE=20115194; PubMed=10648423;
RA   Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
RT   "Novel SH3 protein encoded by the AF3p21 gene is fused to the mixed
RT   lineage leukemia protein in a therapy-related leukemia with
RT   t(3;11)(p21;q23).";
RL   Blood 95:1066-1068(2000).
RN   [11]
RP   INTERACTION WITH SBF1.
RX   MEDLINE=98196760; PubMed=9537414;
RA   Cui X., De Vivo I., Slany R., Miyamoto A., Firestein R., Cleary M.L.;
RT   "Association of SET domain and myotubularin-related proteins modulates
RT   growth control.";
RL   Nat. Genet. 18:331-337(1998).
CC   -!- FUNCTION: Possibly acts as a transcriptional regulatory factor.
CC   -!- SUBUNIT: Interacts with SBF1.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q03164-1; Sequence=Displayed;
CC       Name=14P-18B;
CC         IsoId=Q03164-2; Sequence=VSP_006666;
CC   -!- TISSUE SPECIFICITY: Heart, lung, brain and T and B lymphocytes.
CC   -!- DISEASE: Involved in acute leukemias by chromosomal translocations
CC       t(11;19)(q23;p13.3) that involves MLL and MLLT1/ENL;
CC       t(4;11)(q21;q23) that involves MLL and MLLT2/AF4; t(9;11)(p22;q23)
CC       that involves MLL and MLLT3/AF9; t(6;11)(q27;q23) that involves
CC       MLL and MLLT4/AF6; t(11;17)(q23;q21) that involves MLL and
CC       MLLT6/AF17; t(x;11)(q13;q23) that involves MLL and MLLT7/AFX1;
CC       t(10;11)(p12;q23) that involves MLL and MLLT10/AF10;
CC       t(1;11)(q21;q23) that involves MLL and AF1Q; t(11;19)(q23;p13.3)
CC       that involves MLL and ELL; t(11;19)(q23;p23) that involves MLL and
CC       GAS7; and t(3;11)(p21;q23) that involves MLL and AF3p21.
CC   -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC   -!- SIMILARITY: Contains 1 bromodomain.
CC   -!- SIMILARITY: Contains 1 SET domain.
CC   -!- SIMILARITY: Contains 3 PHD-type zinc fingers.
CC   -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC   -!- SIMILARITY: Contains 1 post-SET domain.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/MLL.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04284; AAA58669.1; -.
DR   EMBL; Z69744; CAA93625.1; -.
DR   EMBL; Z69745; CAA93625.1; JOINED.
DR   EMBL; Z69746; CAA93625.1; JOINED.
DR   EMBL; Z69747; CAA93625.1; JOINED.
DR   EMBL; Z69748; CAA93625.1; JOINED.
DR   EMBL; Z69749; CAA93625.1; JOINED.
DR   EMBL; Z69750; CAA93625.1; JOINED.
DR   EMBL; Z69751; CAA93625.1; JOINED.
DR   EMBL; Z69752; CAA93625.1; JOINED.
DR   EMBL; Z69753; CAA93625.1; JOINED.
DR   EMBL; Z69754; CAA93625.1; JOINED.
DR   EMBL; Z69755; CAA93625.1; JOINED.
DR   EMBL; Z69756; CAA93625.1; JOINED.
DR   EMBL; Z69757; CAA93625.1; JOINED.
DR   EMBL; Z69758; CAA93625.1; JOINED.
DR   EMBL; Z69759; CAA93625.1; JOINED.
DR   EMBL; Z69760; CAA93625.1; JOINED.
DR   EMBL; Z69761; CAA93625.1; JOINED.
DR   EMBL; Z69762; CAA93625.1; JOINED.
DR   EMBL; Z69763; CAA93625.1; JOINED.
DR   EMBL; Z69764; CAA93625.1; JOINED.
DR   EMBL; Z69765; CAA93625.1; JOINED.
DR   EMBL; Z69766; CAA93625.1; JOINED.
DR   EMBL; Z69767; CAA93625.1; JOINED.
DR   EMBL; Z69768; CAA93625.1; JOINED.
DR   EMBL; Z69769; CAA93625.1; JOINED.
DR   EMBL; Z69770; CAA93625.1; JOINED.
DR   EMBL; Z69772; CAA93625.1; JOINED.
DR   EMBL; Z69773; CAA93625.1; JOINED.
DR   EMBL; Z69774; CAA93625.1; JOINED.
DR   EMBL; Z69775; CAA93625.1; JOINED.
DR   EMBL; Z69776; CAA93625.1; JOINED.
DR   EMBL; Z69777; CAA93625.1; JOINED.
DR   EMBL; Z69778; CAA93625.1; JOINED.
DR   EMBL; Z69779; CAA93625.1; JOINED.
DR   EMBL; Z69780; CAA93625.1; JOINED.
DR   EMBL; D14540; BAA03407.1; -.
DR   EMBL; L01986; AAA92511.1; -.
DR   EMBL; U04737; AAA18644.1; -.
DR   EMBL; S78570; AAB34770.1; -.
DR   EMBL; X83604; CAA58584.1; -.
DR   EMBL; S66432; AAB28545.1; -.
DR   EMBL; AF231998; AAG26332.2; ALT_TERM.
DR   PIR; A44265; A44265.
DR   TRANSFAC; T02337; -.
DR   Genew; HGNC:7132; MLL.
DR   MIM; 159555; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0007048; P:oncogenesis; TAS.
DR   GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR003616; PostSET.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR002857; Znf_CXXC.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF02178; AT_hook; 2.
DR   Pfam; PF00628; PHD; 3.
DR   Pfam; PF00856; SET; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   SMART; SM00297; BROMO; 1.

  Query Match             4.1%;  Score 158.5;  DB 1;  Length 3969;
  Best Local Similarity   20.4%;  Pred. No. 0.41;
```

```
  Matches  179;  Conservative   97;  Mismatches  313;  Indels  287;  Gaps   40;

Qy          3 PQKHGGGGGGG-----------------SGPSAGSGGGGFGGS-AAVAAATASGGKSGG 43
              |  |||||||                  ||  | || |    |  |||||  |: | ||
Db         13 PGTTGGGGGGGRRGLGGXPRQRVPALLLPPGPPVGGGGPGAPPSPPAVAAAAAAAGSSGA 72

Qy         44 GSCGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTR 103
              |  ||  : ||:|||||:::: ::      | :  |    :  ||
Db         73 GVPGGAAAASAASSSSASSSSSSSSSASSGPALLRVGPGFDAALQ--------------- 117

Qy        104 NLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTG 163
               : | |               ||::|  |:     |      || |           | |
Db        118 -VSAAI-------------GTNLRR--FRAVFGESGGGGGSGEDEQ----------FLG 150

Qy        164 FFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTK 223
              |          |: |          |:|             : ||    |   |   : :
Db        151 F---------GSDEE----------VRV-------------RSPTRSPSVKTSPRKPRGR 178

Qy        224 P--GNFPSLAVSSNE--FEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEE 279
              |  |:  : |: |:   | | |    ||                          : |   :
Db        179 PRSGSDRNSAILSDPSVFSPLNKSETKS----------------------GDKIKKKDS 215

Qy        280 LPARRKRNREDGEKTFVAQMTVFDKNRRL-QLLDGEYEVAMQEMEECP---------ISK 329
                 :|| |     ||          : : :|  |  | ::::::  |          | |
Db        216 KSIEKKRGR---PPTFPGVKIKITHGKDISELPKGNKEDSLKKIKRTPSATFQQATKIKK 272

Qy        330 KRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQE- 388
               ||       ||  |    |  |  ||   :        :   |  : : | :   :: |
Db        273 LRA-------GKLSPLKSKFKTG-KLQIGRKGVQIVRRRGRPPSTERIKTPSGLLINSEL 324

Qy        389 NKPGSVK------PTQTIAVKESLTTDLQTRKE----KDTPNENRQKLRIFYQFLYNNNT 438
               ||  |:      |  |   ||  |   |: :     :  |:  |     |  | |
Db        325 EKPQKVRKDKEGTPPLT---KEDKTVVRQSPRRIKPVRIIPSSKRTDATIAKQLL----Q 377

Qy        439 RQQTEARDDLHCPWCTLNCRKLYSLLKHLK---------------LCHSRFIFNYVYHP- 482
              | :  |:  :       |  ||: : :|:::                  |||  :  | |
Db        378 RAKKGAQKKIEKEAAQLQGRKVKTQVKNIRQFIMPVVSAISSRIIKTPRRFIEDEDYDPP 437

Qy        483 -KGARIDVSINECYDGSYAGNPQ--------------DIHR--QPGFAFSRNG------- 518
               | ||:: : |  :     |: :               |  |   |    | :
Db        438 IKIARLESTPNSRFSAPSCGSSEKSSAASQHSSQMSSDSSRSSSPSVDTSTDSQASEEIQ 497

Qy        519 ----------------PVKRTPITHILVCRPKRTKASMSEF----------LESEDGEVE 552
                             |: ::|      | :|   |    |          |:|
Db        498 VLPEERSDTPEVHPPLPISQSPENESNDRRSRRYSVSERSFGSRTTKKLSTLQS----AP 553

Qy        553 QQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEF-----SDVN 607
              ||:| ||    |        ||:|      |  :  | ||   ||      |     : :
Db        554 QQQTSSSPPPPLLTPPP---PLQP----ASSISDHTP-WLMPPTIPLASPFLPASTAPMQ 605

Qy        608 EGEKEVMK----LWNLHVMKHGFIADNQMNHACMLFVENYGQK-IIKKNLCRNFMLHLVS 662
               | :::     |    :||        : |       | :: :|:| :  ||    ::
Db        606 GKRKSILREPTFRWT--SLKHSRSEPQYFSSA------KYAKEGLIRKPIFDNFRPPPLT 657

Qy        663 MHDFNLISIMSIDKAVTKLR-----------EMQQ--------KLEKGESASPANEEITE 703
               |   | |  |              :| :       :    |: |   | :|
Db        658 PEDVGFASGFSASGTAASARLFSPLHSGTRFDMHKRSPLLRAPRFTPSEAHSRIFESVTL 717

Qy        704 EQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              | |: | |              |||| : :|:|:
Db        718 PSNRTSAGTSS------------SGVSNRKRKRKV 740


RESULT 7
K1CI_HUMAN
ID   K1CI_HUMAN     STANDARD;      PRT;   622 AA.
AC   P35527; Q14665;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Keratin, type I cytoskeletal 9 (Cytokeratin 9) (K9) (CK 9).
GN   KRT9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Foot sole tissue;
RX   MEDLINE=94131202; PubMed=7507869;
RA   Langbein L., Heid H.W., Moll I., Franke W.W.;
RT   "Molecular characterization of the body site-specific human epidermal
RT   cytokeratin 9: cDNA cloning, amino acid sequence, and tissue
RT   specificity of gene expression.";
RL   Differentiation 55:57-72(1993).
RN   [2]
RP   SEQUENCE OF 449-465.
RX   MEDLINE=90267446; PubMed=2140676;
RA   Rosen E.M., Meromsky L., Romero R., Setter E., Goldberg I.;
RT   "Human placenta contains an epithelial scatter protein.";
RL   Biochem. Biophys. Res. Commun. 168:1082-1088(1990).
RN   [3]
RP   VARIANTS EPPK VAL-156 AND PRO-171.
RX   MEDLINE=94274199; PubMed=7516304;
RA   Hennies H.-C., Zehender D., Kunze J., Kuester W., Reis A.;
RT   "Keratin 9 gene mutational heterogeneity in patients with
RT   epidermolytic palmoplantar keratoderma.";
RL   Hum. Genet. 93:649-654(1994).
RN   [4]
RP   VARIANT EPPK SER-160.
RX   MEDLINE=95015968; PubMed=7523529;
RA   Bonifas J.M., Matsumura K., Chen M.A., Berth-Jones J.,
RA   Hutchinson P.E., Zloczower M., Fritsch P.O., Epstein E.H. Jr.;
RT   "Mutations of keratin 9 in two families with palmoplantar
RT   epidermolytic hyperkeratosis.";
RL   J. Invest. Dermatol. 103:474-477(1994).
RN   [5]
RP   VARIANT EPPK TYR-160.
RX   MEDLINE=94184355; PubMed=7511021;
RA   Torchard D., Blanchet-Bardon C., Serova O., Langbein L., Narod S.,
RA   Janin N., Goguel A.F., Bernheim A., Franke W.W., Lenoir G.M.,
RA   Feunteun J.;
RT   "Epidermolytic palmoplantar keratoderma cosegregates with a keratin 9
RT   mutation in a pedigree with breast and ovarian cancer.";
RL   Nat. Genet. 6:106-110(1994).
RN   [6]
RP   VARIANTS EPPK LYS-160; GLN-162 AND TRP-162.
RX   MEDLINE=94214498; PubMed=7512862;
RA   Reis A., Hennies H.-C., Langbein L., Digweed M., Mischke D.,
RA   Dreschler M., Schroek E., Royer-Pokora B., Franke W.W., Sperling K.,
RA   Kuester W.;
RT   "Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma
RT   (EPPK).";
RL   Nat. Genet. 6:174-179(1994).
RN   [7]
RP   VARIANTS EPPK TRP-162 AND SER-167.
RX   MEDLINE=95164983; PubMed=7532199;
RA   Rothnagel J.A., Wojcik S., Liefer K.M., Dominey A.M., Huber M.,
RA   Hohl D., Roop D.R.;
RT   "Mutations in the 1A domain of keratin 9 in patients with
RT   epidermolytic palmoplantar keratoderma.";
RL   J. Invest. Dermatol. 104:430-433(1995).
RN   [8]
RP   VARIANT EPPK VAL-159.
RX   MEDLINE=97348961; PubMed=9204965;
RA   Endo H., Hatamochi A., Shinkai H.;
RT   "A novel mutation of a leucine residue in coil 1A of keratin 9 in
RT   epidermolytic palmoplantar keratoderma.";
RL   J. Invest. Dermatol. 109:113-115(1997).
RN   [9]
RP   VARIANTS EPPK THR-156; VAL-156 AND GLN-162.
RX   MEDLINE=99072662; PubMed=9856842;
RA   Covello S.P., Irvine A.D., McKenna K.E., Munro C.S., Nevin N.C.,
RA   Smith F.J.D., Uitto J., McLean W.H.I.;
RT   "Mutations in keratin K9 in kindreds with epidermolytic palmoplantar
RT   keratoderma and epidemiology in Northern Ireland.";
```

```
RL   J. Invest. Dermatol. 111:1207-1209(1998).
CC   -!- FUNCTION: MAY SERVE AN IMPORTANT SPECIAL FUNCTION EITHER IN THE
CC       MATURE PALMAR AND PLANTAR SKIN TISSUE OR IN THE MORPHOGENIC
CC       PROGRAM OF THE FORMATION OF THESE TISSUE.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE TERMINALLY DIFFERENTIATED
CC       EPIDERMIS OF PALMS AND SOLES.
CC   -!- DISEASE: DEFECTS IN KRT9 ARE A CAUSE OF EPIDERMOLYTIC PALMOPLANTAR
CC       KERATODERMA (EPPK), AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY
CC       DIFFUSE THICKENING OF THE EPIDERMIS ON THE ENTIRE SURFACE OF
CC       PALMS AND SOLES SHARPLY BORDERED WITH ERYTHEMATOUS MARGINS.
CC   -!- MISCELLANEOUS: There are two types of cytoskeletal and
CC       microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
CC       55 and 56-70 kDa, respectively).
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE A 60 kDa CHAIN OF
CC       PLACENTAL SCATTER PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75015; CAA52924.1; -.
DR   EMBL; Z29074; CAA82315.1; -.
DR   EMBL; S69510; AAC60619.1; -.
DR   PIR; I37984; I37984.
DR   Genew; HGNC:6447; KRT9.
DR   MIM; 144200; -.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR   GO; GO:0008544; P:epidermal differentiation; TAS.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR002957; Keratin_I.
DR   Pfam; PF00038; filament; 1.
DR   PRINTS; PR01248; TYPE1KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Keratin; Disease mutation.
FT   DOMAIN        1    151       HEAD.
FT   DOMAIN      152    460       ROD.
FT   DOMAIN      461    622       TAIL.
FT   DOMAIN      152    187       COIL 1A.
FT   DOMAIN      188    206       LINKER 1.
FT   DOMAIN      207    298       COIL 1B.
FT   DOMAIN      299    321       LINKER 12.
FT   DOMAIN      322    460       COIL 2.
FT   DOMAIN       14     20       POLY-GLY.
FT   VARIANT     156    156       M -> T (IN EPPK).
FT                                /FTId=VAR_010499.
FT   VARIANT     156    156       M -> V (IN EPPK).
FT                                /FTId=VAR_010500.
FT   VARIANT     159    159       L -> V (IN EPPK).
FT                                /FTId=VAR_010501.
FT   VARIANT     160    160       N -> K (IN EPPK).
FT                                /FTId=VAR_003822.
FT   VARIANT     160    160       N -> S (IN EPPK).
FT                                /FTId=VAR_010502.
FT   VARIANT     160    160       N -> Y (IN EPPK).
FT                                /FTId=VAR_010503.
FT   VARIANT     162    162       R -> Q (IN EPPK).
FT                                /FTId=VAR_003823.
FT   VARIANT     162    162       R -> W (IN EPPK).
FT                                /FTId=VAR_003824.
FT   VARIANT     167    167       L -> S (IN EPPK).
FT                                /FTId=VAR_003825.
FT   VARIANT     171    171       Q -> P (IN EPPK).
FT                                /FTId=VAR_010504.
FT   CONFLICT     12     12       T -> SR (IN REF. 1; CAA82315).
SQ   SEQUENCE   622 AA;  61987 MW;  898C3825D4B5ED94 CRC64;

  Query Match             4.0%;  Score 157;  DB 1;  Length 622;
  Best Local Similarity   20.9%;  Pred. No. 0.05;
  Matches   89;  Conservative   63;  Mismatches  165;  Indels  108;  Gaps   19;

Qy          9 GGGGGSGPSAGSGGGGFGGSAAVAAATASGGKS---------GGGSCGG-GGSYSA---- 54
              ||| | | || | |||||| :      : || |         |||| || || | :
Db         73 GGGSGGGFSASSLGGGFGGGSRGFGGASGGGYSSSGGFGGGFGGGSGGGFGGGYGSGFGG 132

Qy         55 --SSSSSAAAAAGAAVLPVKKPKMEHVQADHELFL---QAFEK---------------- 92
                    |    |  :   :|  |: : :    :|   || |:
Db        133 LGGFGGGAGGGDGGILTANEKSTMQELNSRLASYLDKVQALEEANNDLENKIQDWYDKKG 192

Qy         93 PTQIYR----FLRT----RNLIAPIFLHRTLTYMSHRNSRTNIK--RKTFKVDDMLSK-- 140
              |  | :    :  |    :: |  : :    | :   |:|  :   |  |::: | :
Db        193 PAAIQKNYSPYYNTIDDLKDQIVDLTVGNNKTLLDIDNTRMTLDDFRIKFEMEQNLRQGV 252

Qy        141 --------------VEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLE 185
                           :||   | :  :|   |          || |   :   |||
Db        253 DADINGLRQVLDNLTMEKSDLEMQYETLQEELMA-----LKKNHKEEMSQLTGQNS---- 303

Qy        186 VLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMV 245
                        ||:  |   | ||     : |: |     :  | :: |  :  | :
Db        304 ----------GDVNVEINVAP-GKDLTKTLNDMRQ----EYEQL-IAKNRKDIENQYET 346

Qy        246 KSYSLLFRVTRPGR------REFNGMINGETNENIDVNEEL--PARRKRNREDGEKTFVA 297
              :   :   |: |:      :|   : :|     |::  :|   |  ::: || :  :
Db        347 QITQIEHEVSSSGQEVQSSAKEVTQLRHGVQELEIELQSQLSKKAALEKSLEDTKNRYCG 406

Qy        298 QMTVFDK---NRRLQLLDGEYEVAMQEME-------ECPISKKRATWETILDGKRLPPFE 347
              |: :  :   |  |: |  |:  |  |       :  : |:  |:  :|:| :   ||
Db        407 QLQMIQEQISNLEAQITDVRQEIECQNQEYSLLLSIKMRLEKEIETYHNLLEGGQ-EDFE 465

Qy        348 TFSQG 352
              :  |
Db        466 SSGAG 470


RESULT 8
CHD1_MOUSE
ID   CHD1_MOUSE     STANDARD;      PRT;  1711 AA.
AC   P40201;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
GN   CHD1 OR CHD-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93211972; PubMed=8460153;
RA   Delmas V., Stokes D.G., Perry R.P.;
RT   "A mammalian DNA-binding protein that contains a chromodomain and an
RT   SNF2/SWI2-like helicase domain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:2414-2418(1993).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC       IMPORTANT ROLE IN GENE REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: ABUNDANCE IS HIGHER IN CELLS REPRESENTING
CC       EARLY STAGES OF THE B LYMPHOID LINEAGE SUCH AS PRE-B AND B CELLS,
CC       THAN IN CELLS REPRESENTING MATURE PLASMACYTES OR OTHER CELL
CC       LINEAGES SUCH AS FIBROBLASTS.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: Contains 2 chromo domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; L10410; AAB08486.1; -.
DR   HSSP; P23197; 1AP0.
DR   MGD; MGI:88393; Chd1.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003682; F:chromatin binding activity; NAS.
DR   GO; GO:0006333; P:chromatin assembly/disassembly; NAS.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00385; chromo; 2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN        1     70       SER-RICH.
FT   DOMAIN      116    136       SER-RICH.
FT   DOMAIN      270    362       CHROMO 1.
FT   DOMAIN      387    450       CHROMO 2.
FT   NP_BIND     504    511       ATP (POTENTIAL).
FT   SITE        612    615       DEAH BOX.
FT   DOMAIN     1629   1645       3 X 5 AA REPEATS OF H-S-D-H-R.
FT   REPEAT     1629   1633       1.
FT   REPEAT     1635   1639       2.
FT   REPEAT     1641   1645       3.
SQ   SEQUENCE   1711 AA;  196409 MW;  FE3F8F8D13E32E24 CRC64;

  Query Match             4.0%;  Score 153.5;  DB 1;  Length 1711;
  Best Local Similarity   20.1%;  Pred. No. 0.28;
  Matches  124;  Conservative   80;  Mismatches  207;  Indels  205;  Gaps   31;

Qy          5 KHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAA 64
              ::| |    ||   ||  |   ||:: :::  |  :| | |    || | | | | :  :
Db         11 RNGSGESSQSGDDCGSASGSGSGSSSGSSSDGSSSQS-GSSDSDSGSDSGSQSESESDTS 69

Qy         65 GAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSR 124
                :  | ||::  :     | ::      : |    |          |  :  |
Db         70 RENKVQAKPPKVDGAE-----FWKSSPSILAVQRSAMLRKQPQQAQQQRPASSNSGSEED 124

Qy        125 TNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPS---PNSENEQNS 181
              ::    :  || |  :: |   |   :|             :  ||    :||:|:
Db        125 SSSSEDS---DDSSSGAKRKKHNDEDWQMSG-----------SGSPSQLGSDSESEE-- 167

Qy        182 VTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSN 241
                        :|   ||                                : :::|| |
Db        168 ------------ERDKSSCD----------------------------GTESDYEPKN 185

Qy        242 SHMVKSYSLLFRVTRPGRR--EFNG-MINGETNENIDV----------NEELPARRKRN- 287
                      |  :|  |    ||  | |:    ||           |::  :||:
Db        186 K---------VRSRKPQNRSKSKNGKKILGQKKRQIDSSEDEDDEDYDNDKRSSRRQATV 236

Qy        288 ----REDGE-KTFVAQMTVFDKNRRLQLL--------DGEYEVAMQEMEECPISKKRATW 334
                  :|| | ||         | :  |::          | |:|  :: : :| : :| ||
Db        237 NVSYKEDEEMKT--------DSDDLLEVCGEDVPQPEDEEFE-TIERVMDCRVGRKGATG 287

Qy        335 ETI------LDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQE 388
               |        ||     ||   :   :|: ::| | ::  :|              |:| |:
Db        288 ATTTIYAVEADGDPNAGFERNKEPGDIQYLIKWKGWSHIHNT--------WETEETLKQQ 339

Qy        389 NKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN--ENRQKLRIFYQFLYNNNTRQQTEARD 446
              |  |  |             |   :|:::|    :|     : |   ||     | |  |
Db        340 NVRGMKK-----------LDNYKKKDQETKRWLKNASPEDVEY---YN----CQQELTD 380

Qy        447 DLHCPWCTLNCRKLYSLLKHLKLCHSR---------------FIFNYVYHPKGARIDVS 490
              |||         | | ::: : : ||                 ::      || |
Db        381 DLH---------KQYQIVERI-IAHSNQKSAAGLPDYYCKWQGLPYSECSWEDGALISKK 430

Qy        491 INECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPI--THILVCRPKRTKASMSEFLESED 548
                                                                  |:
Db        431 FQTCID-EY--------------FSRN-QSKTTPFKDCKVLKQRPR--------FV---- 462

Qy        549 GEVEQQRTYSSGHNRL 564
                :::| :|  ||  |
Db        463 -ALKKQPSYIGGHEGL 477


RESULT 9
DSX_DROME
ID   DSX_DROME      STANDARD;      PRT;   549 AA.
AC   P23023; P23022; Q9VHY0;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Doublesex protein.
GN   DSX OR CG11094.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS FEMALE AND MALE).
RC   TISSUE=Larva, and Pupae;
RX   MEDLINE=89168451; PubMed=2493994;
RA   Burtis K.C., Baker B.S.;
RT   "Drosophila doublesex gene controls somatic sexual differentiation by
RT   producing alternatively spliced mRNAs encoding related sex-specific
RT   polypeptides.";
RL   Cell 56:997-1010(1989).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM MALE).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
```

```
RN   [3]
RP   DNA-BINDING.
RX   MEDLINE=91330881; PubMed=1907913;
RA   Burtis K.C., Coschigano K.T., Baker B.S., Wensink P.C.;
RT   "The doublesex proteins of Drosophila melanogaster bind directly to a
RT   sex-specific yolk protein gene enhancer.";
RL   EMBO J. 10:2577-2582(1991).
RN   [4]
RP   DNA-BINDING DOMAIN, AND MUTAGENESIS.
RX   MEDLINE=93178426; PubMed=8440242;
RA   Erdman S.E., Burtis K.C.;
RT   "The Drosophila doublesex proteins share a novel zinc finger related
RT   DNA binding domain.";
RL   EMBO J. 12:527-535(1993).
CC   -!- FUNCTION: CONTROLS SOMATIC SEXUAL DIFFERENTIATION. BINDS DIRECTLY
CC       AND SPECIFICALLY TO THE ENHANCER FBE (FAT BODY ENHANCER) OF THE
CC       YOLK PROTEIN 1 AND 2 GENES (YP1 AND YP2). THIS ENHANCER IS
CC       SUFFICIENT TO DIRECT THE FEMALE-SPECIFIC TRANSCRIPTION
CC       CHARACTERISTIC OF THE YP GENES IN ADULT FAT BODIES.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Male;
CC         IsoId=P23023-1; Sequence=Displayed;
CC       Name=Female;
CC         IsoId=P23023-2; Sequence=VSP_001321, VSP_001322;
CC   -!- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.
CC   -!- SIMILARITY: Contains 1 DM domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M25292; AAA17840.1; -.
DR   EMBL; M25293; AAA17841.1; -.
DR   EMBL; M25294; AAA17842.1; -.
DR   EMBL; AE003676; AAF54168.1; -.
DR   PIR; A32372; A32372.
DR   PIR; B32372; B32372.
DR   PDB; 1LPV; 02-OCT-02.
DR   TRANSFAC; T00955; -.
DR   TRANSFAC; T00956; -.
DR   FlyBase; FBgn0000504; dsx.
DR   GO; GO:0003729; F:mRNA binding activity; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007619; P:courtship behavior; NAS.
DR   GO; GO:0045497; P:female analia morphogenesis (sensu Holometa. . .; NAS.
DR   GO; GO:0007486; P:female genital morphogenesis (sensu Holomet. . .; IGI.
DR   GO; GO:0019101; P:female somatic sex determination; NAS.
DR   GO; GO:0007483; P:genital disc metamorphosis; NAS.
DR   GO; GO:0045496; P:male analia morphogenesis (sensu Holometabola); NAS.
DR   GO; GO:0045433; P:male courtship behavior (sensu Drosophila),. . .; NAS.
DR   GO; GO:0007485; P:male genital morphogenesis (sensu Holometab. . .; IGI.
DR   GO; GO:0019102; P:male somatic sex determination; NAS.
DR   GO; GO:0007548; P:sex differentiation; NAS.
DR   InterPro; IPR001275; DM_DNA-binding.
DR   Pfam; PF00751; DM-domain; 1.
DR   SMART; SM00301; DM; 1.
DR   PROSITE; PS40000; DM_DOMAIN_1; 1.
DR   PROSITE; PS50809; DM_DOMAIN_2; 1.
KW   Sexual differentiation; Alternative splicing; DNA-binding;
KW   Transcription regulation; Nuclear protein; Zinc; Metal-binding;
KW   3D-structure.
FT   DNA_BIND     44     91       DM.
FT   DOMAIN      119    224       HIS-RICH.
FT   DOMAIN      267    296       SER/GLY-RICH.
FT   VARSPLIC    398    427       ARVEINRTVAQIYYNYYTPMALVNGAPMYL -> GQYVVNE
FT                                YSRQHNLNIYDGGELRNTTRQCG (in isoform
FT                                Female).
FT                                /FTId=VSP_001321.
FT   VARSPLIC    428    549       Missing (in isoform Female).
FT                                /FTId=VSP_001322.
FT   MUTAGEN      47     47       C->A,H: ABOLISHES DNA-BINDING.
FT   MUTAGEN      50     50       H->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      59     59       H->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      68     68       C->D,Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      70     70       C->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      91     91       R->Q: ABOLISHES DNA-BINDING.
SQ   SEQUENCE   549 AA;  57409 MW;  3C1B92724E4CE083 CRC64;

  Query Match             3.9%;  Score 153;  DB 1;  Length 549;
  Best Local Similarity   27.5%;  Pred. No. 0.073;
  Matches   47;  Conservative   27;  Mismatches   47;  Indels   50;  Gaps    5;

Qy         11 GGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSA---------------- 54
              ||  ||:  | |||   |:  ||||:| | ||||  |||||
Db        242 GGSVGPATSSSGGGAPSSSNAAAATSSNGSSGGGGGGGGGSSGGGAGGGRSSGTSVITSA 301

Qy         55 ----------------------SSSSSAAAAAGAAVLP----VKKPKMEHVQADHELFLQAFE 91
                                    ||| | :: :|||:||    | :   :|    ::|| :
Db        302 DHHMTTVPTPAQSLEGSCDSSSPSPSSTSGAAILPISVSVNRKNGANVPLGQDVFLDYCQ 361

Qy         92 KPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVE 142
              |  : :|:        |  |   | |:  :::  ||:  : ::::  ::||
Db        362 KLLEKFRY--------PWEL-MPLMYVILKDADANIEEASRRIEE--ARVE 401


RESULT 10
K2C1_MOUSE
ID   K2C1_MOUSE     STANDARD;      PRT;   627 AA.
AC   P04104;
DT   01-NOV-1986 (Rel. 03, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Keratin, type II cytoskeletal 1 (Cytokeratin 1) (67 kDa cytokeratin).
GN   KRT1 OR KRT2-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 83-628 FROM N.A.
RX   MEDLINE=85207740; PubMed=2581964;
RA   Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA   Roop D.R.;
RT   "Amino acid sequences of mouse and human epidermal type II keratins
RT   of Mr 67,000 provide a systematic basis for the structural and
RT   functional diversity of the end domains of keratin intermediate
RT   filament subunits.";
RL   J. Biol. Chem. 260:7142-7149(1985).
RN   [2]
RP   REVISIONS, AND SEQUENCE FROM N.A.
RA   Roop D.R.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC       MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC       (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M10937; AAD05191.1; -.
DR   SWISS-2DPAGE; P04104; MOUSE.
DR   MGD; MGI:96698; Krt2-1.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR003054; Keratin_II.
```

```
DR   Pfam; PF00038; filament; 1.
DR   PRINTS; PR01276; TYPE2KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Keratin.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    186       HEAD.
FT   DOMAIN      187    496       ROD.
FT   DOMAIN      497    627       TAIL.
FT   DOMAIN      187    222       COIL 1A.
FT   DOMAIN      223    242       LINKER 1.
FT   DOMAIN      243    333       COIL 1B.
FT   DOMAIN      334    357       LINKER 12.
FT   DOMAIN      358    496       COIL 2.
FT   SITE        451    451       STUTTER.
SQ   SEQUENCE   627 AA;  65092 MW;  EF7E848654539578 CRC64;

  Query Match             3.9%;  Score 153;  DB 1;  Length 627;
  Best Local Similarity   22.0%;  Pred. No. 0.086;
  Matches   81;  Conservative   53;  Mismatches  140;  Indels   94;  Gaps   15;

Qy          6 HGGGG-GGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAA 64
              :|||| ||||    | || |||| :       |||  |||  || |   : |
Db        106 YGGGGFGGGSFGGGGFGGSGFGGGSGGGGGFGSGGGFGGGRFGGYGPVCSPSGIQEVTLN 165

Qy         65 GAAVLP-----------VKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPI---F 110
               : : |           ||  : | ::: :: |    :|  .  |||  :| :
Db        166 QSLLQPLNVKVDPQIQKVKSQEREQIKSLNDKFASFIDK----VRFLEQQNKVLQTKWEL 221

Qy        111 LHRTLTYMSHRN------SRTNIKRKTFKVDDMLSKVEKMKGE-QESHSLSAHLQLTFTG 163
              | :  |    :|      :  :| |:  ||| : |   :|: | :    |    :  :
Db        222 LQQVDTTTRTQNLDPFFENYISILRR--KVDSLKSDQSRMESELKNMQDLVEEYRTKYED 279

Qy        164 FFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDV-SCPIRQVPTGKKQVPLIPDLNQT 222
              :|     |:|||   ||:           :||| |  : :|    |:  |  |:|
Db        280 EINKR----TNAENE--FVTI----------KKDVDSAYMTKVELQAKRDALQQDIN-- 320

Qy        223 KPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVN----- 277
                  |  ::   | |   |            | | ::: : |   |::
Db        321 ---------FFSTLYQMEMSQMQTQIS----------ETNVVLSMDNNRQFDLDGIISE 360

Qy        278 -----EELPARRKRNRE---------------DGEKTFVAQMTVFDKNRRLQLLDGEYE 316
                  : :  | |   |                |:    :| : : || :| |  | :
Db        361 VKAQYDSICQRSKAEAETFYQSKYEELQITAGKHGDSVRNTKMEISELNRMIQRLRSEID 420

Qy        317 VAMQEMEE 324
                ::: :
Db        421 GCKKQISQ 428


RESULT 11
IFE_BRALA
ID   IFE_BRALA      STANDARD;      PRT;   464 AA.
AC   Q04948;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Non-neuronal cytoplasmic intermediate filament protein.
OS   Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7740;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92354564; PubMed=1644059;
RA   Riemer D., Dodemont H., Weber K.;
RT   "Analysis of the cDNA and gene encoding a cytoplasmic intermediate
RT   filament (IF) protein from the cephalochordate Branchiostoma
RT   lanceolatum; implications for the evolution of the IF protein
RT   family.";
RL   Eur. J. Cell Biol. 58:128-135(1992).
CC   -!- SUBUNIT: CAN FORM HOMOPOLYMERS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64522; CAA45827.1; -.
DR   PIR; A56600; A56600.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR002957; Keratin_I.
DR   Pfam; PF00038; filament; 1.
DR   PRINTS; PR01248; TYPE1KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil.
FT   DOMAIN        1    101       HEAD.
FT   DOMAIN      102    413       ROD.
FT   DOMAIN      414    464       TAIL.
FT   DOMAIN      102    133       COIL 1A.
FT   DOMAIN      134    144       LINKER 1.
FT   DOMAIN      145    237       COIL 1B.
FT   DOMAIN      238    264       LINKER 2.
FT   DOMAIN      265    413       COIL 2.
SQ   SEQUENCE   464 AA;  51300 MW;  095CB4025A761608 CRC64;

  Query Match             3.9%;  Score 151;  DB 1;  Length 464;
  Best Local Similarity   19.8%;  Pred. No. 0.078;
  Matches   95;  Conservative   77;  Mismatches  158;  Indels  150;  Gaps   20;

Qy          7 GGGGGGGSGPSAGSGGGGF-GGSAAVAAATASGGKSGGG--SCGGGGSYSASSSSSAAAA 63
              ||||||| | :: |  | | ||   : |   :  :| |   | ||||  :|  ::   |:
Db         18 GGGGGGGGGRASYSSSGRFSGGGGRMRAGGVTSRRSVGSSYSAGGGGRVNAVMAAGVMAS 77

Qy         64 AG---AAVLPVKKPKMEHVQADHEL------FLQAFEKPTQIYRFLRTRN--LIAPIFLH 112
                |    |:  :   :|      ||      |   :|     |:|: ||  | | | :
Db         78 VGGPAGALQTLSDARMTRAHEKQELSHLNDRFASYIDK----VRYLQERNSKLEAQIKIQ 133

Qy        113 RT--------------------LTYMSHRNSRTNIKRKTFKV----------------DD 136
              :                    :  :|:  ::  |:|  ::                 :
Db        134 ESREAPNIKDLYEKELRDLRALVDELSNDKAQLEIERNNWQEQAEDYKMKWQDEAGLRSE 193

Qy        137 MLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
              : :::|::| | :: :: | | |                 ||: ::   |:  :|  | :
Db        194 LEAEIERLKKELDAATM-ARLDL---------------ENKLSTAQEEIDFLKRVHDEE 236

Qy        197 KDVSCPIRQVPTGKKQVPLIPDLNQTKPGNF---PSLAVSSNEFEPSNSHMVKSYSLLFR 253
                  |||:      :   | :::       |   | |  :  |             | | |
Db        237 ------IRQLQDQLNESLTIVEVDSRAASTFAPGPDLTEALREIR-------TQYEELGR 283

Qy        254 VTRPGRREFNGMINGETNENIDVN-----EELPARRKRNREDGEKTFVAQMTVFDKNR-- 306
              | |               |: ||       ||  :|:|   | |    |:  | :  |
Db        284 VNR--------------EDADVKYKQKFSELAHQRER---DNEALMTARTEVTELRRNL 325

Qy        307 -------------------------RLQLLDGEYEVAMQEMEE------CPISKKRATW 334
                                      |:||   ||: |::::|:        ::::   :
Db        326 NSLVAENEALKSKNHALEGSLAELEARMQLEIEEYQAAIRDLEQELETKTSEMAQQMQAY 385

Qy        335 ETILDGKRLPPFETFSQGPTLQ-FTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGS 393
              : ::| |     |  :   |:  :|  ||: :       :  :: :| |     | ||
Db        386 KMLMDTKMALDMEIAAYRKLLEGEEIRLFGESKEG-----VQQTSSSSSSSYQYSMKSGS 440


RESULT 12
TRK1_YEAST
ID   TRK1_YEAST     STANDARD;      PRT;  1235 AA.
AC   P12685;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Potassium transport protein, high-affinity.
GN   TRK1 OR YJL129C OR J0693.
```

```
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88302204; PubMed=3043197;
RA   Gaber R.F., Styles C.A., Fink G.R.;
RT   "TRK1 encodes a plasma membrane protein required for high-affinity
RT   potassium transport in Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 8:2848-2859(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=97103775; PubMed=8948101;
RA   Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT   "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT   reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT   SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,
RT   three remnant delta elements and a Ty4 transposon.";
RL   Yeast 12:1471-1474(1996).
CC   -!- FUNCTION: THIS PROTEIN IS REQUIRED FOR HIGH-AFFINITY POTASSIUM
CC       TRANSPORT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21328; AAA34728.1; -.
DR   EMBL; Z49404; CAA89424.1; -.
DR   PIR; S05849; PWBYH.
DR   SGD; S0003665; TRK1.
DR   GO; GO:0015079; F:potassium ion transporter activity; IDA.
DR   GO; GO:0030007; P:potassium ion homeostasis; IDA.
DR   InterPro; IPR003445; Cat_transpt.
DR   InterPro; IPR004773; Ktransp_euk.
DR   Pfam; PF02386; TrkH; 1.
DR   TIGRFAMs; TIGR00934; 2a38euk; 1.
KW   Transport; Transmembrane; Potassium transport; Glycoprotein.
FT   TRANSMEM     49     70       POTENTIAL.
FT   TRANSMEM     78     98       POTENTIAL.
FT   TRANSMEM    107    127       POTENTIAL.
FT   TRANSMEM    778    800       POTENTIAL.
FT   TRANSMEM    813    834       POTENTIAL.
FT   TRANSMEM    838    858       POTENTIAL.
FT   TRANSMEM    862    882       POTENTIAL.
FT   TRANSMEM    898    918       POTENTIAL.
FT   TRANSMEM    923    943       POTENTIAL.
FT   TRANSMEM    971    991       POTENTIAL.
FT   TRANSMEM   1078   1098       POTENTIAL.
FT   TRANSMEM   1111   1131       POTENTIAL.
FT   CARBOHYD    100    100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    251    251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    383    383       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1030   1030       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1135   1135       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1235 AA;  141072 MW;  BCE9FDBA0BA0982B CRC64;
```

```
  Query Match             3.8%;  Score 146.5;  DB 1;  Length 1235;
  Best Local Similarity   20.1%;  Pred. No. 0.47;
  Matches  154;  Conservative   90;  Mismatches  227;  Indels  295;  Gaps   39;

Qy         97 YRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKV----------EKMK 145
              ::  ||: ::      ||:|       ||:  ||  | ||   |:          : :
Db        148 FKMRRTKTILERELTARTMTKNRTGTQRTSYPRKQAKTDDFQEKLFSGEMVNRDEQDSVH 207

Qy        146 GEQESHSLSAHLQLTFTGF---------FHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
              :| || :|       |           | | |:   | | ::|:
Db        208 SDQNSHDISRDSSNNNTNHNGSSGSLDDFVKEDETDDNGEYQENN-------------- 252

Qy        197 KDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEF-EPSNSHMVKSYSLLF--- 252
                                          :: ::  |||   : | :   |  || |
Db        253 ----------------------------SYSTVGSSSNTVADESLNQKPKPSSLRFDEP 283

Qy        253 -RVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLL 311
                 || |               |  |  |:|: :|:         ::  |  | : :|
Db        284 HSKQRPAR---------------VPSEKFAKRRGSRD---------ISPADMYRSIMML 318

Qy        312 DGEYEVAMQEMEECPI---------------SKKRATWETILDGKRLPPFETFSQGPTL 355
              |::| |  | |  |:                  |:||    ::| ::   :  ::  |
Db        319 QGKHE-ATAEDEGPPLVIGSPADGTRYKSNVNKLKKATG---INGNKIKIRDKGNESNTD 374

Qy        356 QFTLRWTGETNDKSTAPIAKPLATRNSESLHQE--NKPGSVKPTQT-------IAVKESL 406
              | ::  : | |  ||| ::      :  |||    ||  |:: | |       ||: ::
Db        375 QNSV--SSEAN--STASVS------DESSLHTNFGNKVPSLR-TNTHRSNSGPIAITDNA 423

Qy        407 TTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTR---QQTEARDDLHCPWCTLNCRKLYS- 462
              ||     :|  |:          ||       |   ::    |||: |  ::  ||  |
Db        424 ETD-----KKHGPS---------IQFDITKPPRKISKRVSTFDDLN-PKSSVLYRKKASK 468

Qy        463 --LLKHLKLCHSRFIFNYVYHPKGARIDVSI-NECYDGSYAGNPQDIHRQPGFAFSRNGP 519
                |:||               ||  ||   |      ||   |            | |
Db        469 KYLMKHF-------------PKARRIRQQIKRRLSTGSIEKN-----------SSNNV 502

Qy        520 VKRTPITHILVCRPKRTKASMSE--------------FLESEDGEVEQQRTYSSGHNRL 564
              | |||               | :                | ::| |: | :|   |
Db        503 SDRKPIT------------DMDDDDDDDDNDGDNNEEYFADNESGD-EDERVQQSEP--- 546

Qy        565 YFHSDTCLPLRPQEME---------------VDSEDEKDPEWLREKTITQIEEFSDVNE 608
              |||: |    |: |                |    :     | :||  : |  |:||
Db        547 --HSDSELKSHQQQQEKHQLQQNLHRMYKTKSFDDNRSRAVPMERSRTI-DMAEAKDLNE 603

Qy        609 ------GEKEVMKLWNLH---------------VMKHG-FIADNQMNHACMLFVENYGQK 646
                    :| | : |  |                : :|| : :|:  |:       | |
Db        604 LARTPDFQKMVYQNWKAHHRKKPNFRKRGWNNKIFEHGPYASDSDRNYPDN---SNTG-- 658

Qy        647 IIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQN 706
                  | :||         ||:  | :          |    |:::| :|| |
Db        659 --------NSILHYAE-------SILHHDGS--------HKNGSEEASSDSNENIYSTNG 695

Qy        707 GT----ANGFSEINSKEKA---------LETDSVSGVSKQSKKQKL 739
              |:      | :   |  |:           : |    :|| | |  |
Db        696 GSDHNGLNNYPTYNDDEEGYYGLHFDTDYDLDPRHDLSKGSGKTYL 741


RESULT 13
MSP1_PLAFW
ID   MSP1_PLAFW     STANDARD;      PRT;  1639 AA.
AC   P04933;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Wellcome).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5848;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86014355; PubMed=2995820;
```

```
RA   Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA   Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA   Freeman R.R.;
RT   "Primary structure of the precursor to the three major surface
RT   antigens of Plasmodium falciparum merozoites.";
RL   Nature 317:270-273(1985).
RN   [2]
RP   REVISIONS.
RA   Holder A.A.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC       kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02919; CAA26676.1; -.
DR   PIR; A24594; A24594.
DR   PIR; S05603; S05603.
DR   PDB; 1CEJ; 28-MAY-99.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor; 3D-structure.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1639       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    768    768       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    920    920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1058   1058       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1174   1174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1445   1445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1526   1526       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1639 AA;  187618 MW;  2C255B6616C87F6E CRC64;

  Query Match             3.8%;  Score 146.5;  DB 1;  Length 1639;
  Best Local Similarity   20.6%;  Pred. No. 0.67;
  Matches  168;  Conservative  123;  Mismatches  291;  Indels  235;  Gaps   46;

Qy         10 GGGGSGPSAGSGGG-GFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAV 68
              |  ||  | ||||    ||| |   : ||||    |  |     : | :|| : |   |
Db         68 GSKGSVASGGSGGSVASGGSVASGGSVASGGSVASGGSGNSRRTNPSDNSSDSDAKSYA- 126

Qy         69 LPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIK 128
                  ::|   :: | ::  : | |::    | :::       ||          ||
Db        127 ------DLKHRVRNYLLTIKELKYP-QLFDL--TNHML-------TLC--------DNIH 162

Qy        129 RKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLL 188
                 : :|          | :| : |   |   |                     |   |
Db        163 GFKYLID----------GYEEINELLYKLNFYFD--------------------LLRAKL 192

Qy        189 VKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSY 248
                 ||     : |           |:|             :| : :|| :     ::|
Db        193 NDVC----ANDYC----------QIPF-------------NLKIRANELD-----VLK-- 218

Qy        249 SLLFRVTRP--------GRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
               |:|   :|       |: |     | :| |||  || :        |: :||
Db        219 KLVFGYRKPLDNIKDNVGKMEDYIKKNKKTIENI--NELI--------EESKKTIDKNKN 268

Qy        301 VFDKNRRLQLLDGEYEVAM--QEMEEC----------PISKKRATWETILDG----KRL 343
                 :  : :|   :|::::  :::||                ||    : :||     |
Db        269 ATKEEEKKKLYQAQYDLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNP 328

Qy        344 PPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSES----------LHQENKPG 392
             ||  :    |          | ::|    ||| :   | :|           | ::||
Db        329 PPANS-GNTPNTLLDKNKKIEEHEKEIKEIAKTIKF-NIDSLFTDPLELEYYLREKNK-- 384

Qy        393 SVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN--NNTRQQTEARDDLHC 450
                  | :   : |     ||   |||       : |   ||  ||   :  :  ||
Db        385 ------NIDISAKVET-----KESTEPNEYPN--GVTYPLSYNDINNALNELNSFGDLIN 431

Qy        451 PW----------CTLNCRKLY--SLLKHLKLCHSRFIFN-YVYHPKGARI-DVS------ 490
             |:           | | || :   : : :|:   :   :   |  :   : |::
Db        432 PFDYTKEPSKNIYTDNERKKFINEIKEKIKIEKKKIESDKKSYEDRSKSLNDITKEYEKL 491

Qy        491 INECYDGSYAGNPQ--DIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEF---LE 545
             :|| ||  :  |    :  :  |  :|     |   :||        | ||       ||
Db        492 LNEIYDSKFNNNIDLTNFEKMMGKRYS----YKVEKLTH------HNTFASYENSKHNLE 541

Qy        546 SEDGEVEQQRTYSSGH----NRLYFHSDTCLPLRPQEMEVDSED-EKDPEWLREKTITQI 600
                 ::   ||  :    | :: :    :   |:|   |: :|| | | ||,||:
Db        542 KLTKALKYMEDYSLRNIVVEKELKYYKNLISKIE-NEIETLVENIKKDEEQLFEKKITK- 599

Qy        601 EEFSDVNEGEKEVMKLWNL-HVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLH 659
                | |: ::::::: ::  |      :  |:::       |    ::| ||:     :|
Db        600 ----DENKPDEKILEVSDIVKVQVQKVLLMNKID-------ELKKTQLILKNVELKHNIH 648

Qy        660 LVSMH-------DFNLISIMSIDKAVTKLREMQQKLE----------KGESASPANEEIT 702
             : : :        : ||    : | : ||:    |:|          | |  |  :|  |
Db        649 VPNSYKQENKQEPYYLI---VLKKEIDKLKVFMPKVESLINEEKKNIKTEGQSDNSEPST 705

Qy        703 EEQNGTANGFSEINSKEK---ALETDSVSGVSKQSKK 736
             |   |   | :      ::   ||| |||   ::: |:
Db        706 E---GEITGQATTKPGQQAGSALEGDSVQAQAQEQKQ 739


RESULT 14
GOG4_HUMAN
ID   GOG4_HUMAN     STANDARD;      PRT;  2230 AA.
AC   Q13439; Q13270; Q13654; Q14436;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
DE   golgin) (Golgin-245) (72.1 protein).
GN   GOLGA4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=96215236; PubMed=8626529;
RA   Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT   "Molecular characterization of trans-Golgi p230: a human peripheral
RT   membrane protein encoded by a gene on chromosome 6p12-22 contains
RT   extensive coiled-coil alpha-helical domains and a granin motif.";
RL   J. Biol. Chem. 271:8328-8337(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Seelig H.P.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 131-2230 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=96125112; PubMed=8537393;
RA   Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT   "Molecular characterization of golgin-245, a novel Golgi complex
RT   protein containing a granin signature.";
RL   J. Biol. Chem. 270:31262-31268(1995).
RN   [4]
RP   SEQUENCE OF 524-672 FROM N.A.
```

```
RC   TISSUE=Gastric fundus;
RA   Balague C.;
RL   Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC   -!- FUNCTION: May play a role in vesicular transport from the trans-
CC       Golgi.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
CC       GOLGI MEMBRANE.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q13439-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
CC       Name=3;
CC         IsoId=Q13439-3; Sequence=VSP_004274;
CC       Name=4;
CC         IsoId=Q13439-4; Sequence=VSP_004275;
CC   -!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND
CC       IN HEPATITIS B.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U41740; AAC50434.1; -.
DR   EMBL; X82834; CAA58041.1; -.
DR   EMBL; U31906; AAC51791.1; -.
DR   EMBL; X76942; CAA54261.1; -.
DR   Genew; HGNC:4427; GOLGA4.
DR   MIM; 602509; -.
DR   GO; GO:0005802; C:Golgi trans-face; TAS.
DR   GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR   InterPro; IPR000237; GRIP_domain.
DR   Pfam; PF01465; GRIP; 1.
KW   Golgi stack; Antigen; Coiled coil; Alternative splicing.
FT   DOMAIN      133    237       COILED COIL (POTENTIAL).
FT   DOMAIN      276   1011       COILED COIL (POTENTIAL).
FT   DOMAIN     1033   1214       COILED COIL (POTENTIAL).
FT   DOMAIN     1259   2152       COILED COIL (POTENTIAL).
FT   VARSPLIC   2154   2185       TPYKGGNLYHTDVSLFGEPTEFEYLRKVLFEY -> HLTKV
FT                                AICTIRMSHSLENLPNLSICEKCFLSI (in isoform
FT                                2).
FT                                /FTId=VSP_004272.
FT   VARSPLIC   2186   2230       Missing (in isoform 2).
FT                                /FTId=VSP_004273.
FT   VARSPLIC   2103   2109       Missing (in isoform 3).
FT                                /FTId=VSP_004274.
FT   VARSPLIC   2222   2230       FTSPRSGIF -> SWLRSSS (in isoform 4).
FT                                /FTId=VSP_004275.
FT   CONFLICT    188    188       R -> K (IN REF. 3).
FT   CONFLICT    220    220       Y -> H (IN REF. 3).
FT   CONFLICT    276    276       T -> A (IN REF. 3).
FT   CONFLICT    584    584       K -> E (IN REF. 3).
FT   CONFLICT    628    628       T -> A (IN REF. 3).
FT   CONFLICT    630    630       K -> E (IN REF. 3).
FT   CONFLICT    682    682       K -> N (IN REF. 3).
SQ   SEQUENCE   2230 AA;  261139 MW;  3BB733DB1EA86134 CRC64;

  Query Match             3.8%;  Score 146;  DB 1;  Length 2230;
  Best Local Similarity   19.1%;  Pred. No. 1.1;
  Matches 137;  Conservative  119;  Mismatches  245;  Indels  216;  Gaps   34;

Qy        124 RTNIKRKTFK--VDDMLSKVEKMKGEQESHSLSAHL----------------QLTFTGF 164
              :|   |: |  :|: :  :||   |:|  ||   |                  |:
Db        425 KTEEARRKLKAEMDEQIKTIEK-TSEEERISLQQELSRVKQEVVDVMKKSSEEQIAKLQK 483

Qy        165 FHKNDKPSPNSE----------NEQNSVTLEVL---LVKVCHKKRKDVSCPIRQVPTG 209
               |: :     |            ||  | ||     :|:  :| :  |  : ::
```

```
Db        484 LHEKELARKEQELTKKLQTREREFQEQMKVALEKSQSEYLKISQEKEQQESLALEELELQ 543

Qy        210 KKQV-----PLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNG 264
              || :       : || |             | |   : :::  | | :  :  : :
Db        544 KKAILTESENKLRDLQQ-------------EAETYRTRILELESSLEKSLQENKNQSKD 589

Qy        265 M---INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQE 321
              :   :  | |::   |:|:    :::: : |     |  ::    :||:|  :|:  |::
Db        590 LAVHLEAEKNKH---NKEITVMVEKHKTELESLKHQQDALW--TEKLQVLKQQYQTEMEK 644

Qy        322 M-EECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPI------A 374
              : |:|   |      ||:|  | :  |:   :             | |:|:   :
Db        645 LREKCEQEK-----ETLLKDKEI-IFQAHIE-----------EMNEKTLEKLDVKQTEL 686

Qy        375 KPLATRNSESL---HQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQ 431
              : |::  || |   |:  :  ||   ||  :|: |   :  :|     | ::|::    :
Db        687 ESLSSELSEVLKARHKLEEELSVLKDQTDKMKQELEAKMDEQK-----NHHQQQVDSIIK 741

Qy        432 FLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLL----KHLKLCHSRFIFNYVYHPKGARI 487
                :  | :   :| ::         :|  ||    ||||         :  | :
Db        742 EHEVSIQRTEKALKDQIN---------QLELLLKERDKHLK--------EHQAHVENLEA 784

Qy        488 DVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEF---- 543
              |:  :|  :   |    |: :    |            ||      ::|||   :
Db        785 DIKRSE-GELQQASAKLDVFQSYQSA-----------TH------EQTKAYEEQLAQLQ 825

Qy        544 -----LESED-------GEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDS-----EDE 586
                  ||:|         ||| |:              | |  |   :::|       | :
Db        826 QKLLDLETERILLTKQVAEVEAQK-----------KDVCTELDAHKIQVQDLMQQLEKQ 873

Qy        587 KDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQK 646
                  : |::||: | | : :| ||  :   : | |   |   ||       :|   ||
Db        874 NSEMEQKVKSLTQVYE-SKLEDGNKEQEQTKQILVEKENMIL--QMREGQKKEIEILTQK 930

Qy        647 IIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEE------ 700
              :  |           |:|  |       ::  || :  ::|:|| :  :    :|
Db        931 LSAKE---------DSIHILN-------EEYETKFKNQEKKMEKVKQKAKEMQETLKKKL 974

Qy        701 ------ITEEQNGTANGFSEINSKEKALE-------------TDSVSGVSKQSKKQ 737
                   : :|   ||    |:: |||                 :|:|| :    |:|
Db        975 LDQEAKLKKELENTA---LELSQKEKQFNAKMLEMAQANSAGISDAVSRLETNQKEQ 1028


RESULT 15
K2C1_HUMAN
ID   K2C1_HUMAN     STANDARD;      PRT;   643 AA.
AC   P04264; Q14720; Q9H298;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Keratin, type II cytoskeletal 1 (Cytokeratin 1) (K1) (CK 1) (67 kDa
DE   cytokeratin) (Hair alpha protein).
GN   KRT1 OR KRTA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85166239; PubMed=2580302;
RA   Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;
RT   "Structure of a gene for the human epidermal 67-kDa keratin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Whittock N.V., Eady R.A.J., McGrath J.A.;
RT   "Genomic organization of the human keratin 1 gene.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hatsell S.J., Eady R.A.J., Wennerstrand L., Dopping-Hepenstal P.,
RA   Leigh I.M., Munro C., Kelsell D.P.;
RT   "Novel splice site mutation in keratin 1 underlies mild epidermolytic
```

```
RT   palmoplantar keratoderma in three kindreds.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   PRELIMINARY SEQUENCE OF 151-643 FROM N.A.
RX   MEDLINE=85207740; PubMed=2581964;
RA   Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA   Roop D.R.;
RT   "Amino acid sequences of mouse and human epidermal type II keratins
RT   of Mr 67,000 provide a systematic basis for the structural and
RT   functional diversity of the end domains of keratin intermediate
RT   filament subunits.";
RL   J. Biol. Chem. 260:7142-7149(1985).
RN   [5]
RP   REVISIONS, AND VARIANT EHK PRO-160.
RX   MEDLINE=92386601; PubMed=1381288;
RA   Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,
RA   Compton J.G., Steinert P.M.;
RT   "A leucine-->proline mutation in the H1 subdomain of keratin 1
RT   causes epidermolytic hyperkeratosis.";
RL   Cell 70:821-828(1992).
RN   [6]
RP   VARIANTS EHK GLY-154; SER-187 AND PRO-192.
RX   MEDLINE=94117869; PubMed=7507151;
RA   Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,
RA   Steinert P.M., Compton J.G.;
RT   "Mutations in the H1 and 1A domains in the keratin 1 gene in
RT   epidermolytic hyperkeratosis.";
RL   J. Invest. Dermatol. 102:17-23(1994).
RN   [7]
RP   VARIANTS EHK PRO-185 AND SER-187.
RX   MEDLINE=94117870; PubMed=7507152;
RA   McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J., McMillan J.R.,
RA   Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA   Morley S.M.;
RT   "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT   congenital ichthyosiform erythroderma (BCIE).";
RL   J. Invest. Dermatol. 102:24-30(1994).
RN   [8]
RP   VARIANT EHK GLN-489.
RX   MEDLINE=92376531; PubMed=1380725;
RA   Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA   Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT   "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT   hyperkeratosis.";
RL   Science 257:1128-1130(1992).
RN   [9]
RP   VARIANT ALLELE 1B.
RX   MEDLINE=93107743; PubMed=1281859;
RA   Korge B.P., Compton J.G., Steinert P.M., Mischke D.;
RT   "The two size alleles of human keratin 1 are due to a deletion in the
RT   glycine-rich carboxyl-terminal V2 subdomain.";
RL   J. Invest. Dermatol. 99:697-702(1992).
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC       KERATIN 1 IS GENERALLY ASSOCIATED WITH KERATIN 10.
CC   -!- TISSUE SPECIFICITY: THE SOURCE OF THIS PROTEIN IS NEONATAL
CC       FORESKIN. THE 67-KDA TYPE II KERATINS ARE EXPRESSED IN TERMINALLY
CC       DIFFERENTIATING EPIDERMIS.
CC   -!- POLYMORPHISM: THERE ARE TWO SIZE VARIANTS OF KRT1, TERMED 1A AND
CC       1B WITH ALLELIC FREQUENCIES OF 0.61 AND 0.39. 1B LACKS 7 LACKS
CC       7 RESIDUES COMPARED TO 1A.
CC   -!- DISEASE: DEFECTS IN KRT1 ARE A CAUSE OF EPIDERMOLYTIC
CC       HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC       ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER
CC       CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC       CORNEUM.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC       MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II
CC       (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF237621; AAF60327.1; -.
DR   EMBL; M98776; AAB47721.1; -.
DR   EMBL; AF304164; AAG41947.1; -.
DR   EMBL; M10938; AAA36153.1; ALT_SEQ.
DR   Aarhus/Ghent-2DPAGE; 4606; NEPHGE.
DR   Genew; HGNC:6412; KRT1.
DR   MIM; 139350; -.
DR   MIM; 113800; -.
DR   GO; GO:0005856; C:cytoskeleton; TAS.
DR   GO; GO:0005882; C:intermediate filament; TAS.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR   GO; GO:0008544; P:epidermal differentiation; TAS.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR003054; Keratin_II.
DR   Pfam; PF00038; filament; 1.
DR   PRINTS; PR01276; TYPE2KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Keratin; Disease mutation;
KW   Polymorphism; Phosphorylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    178       HEAD.
FT   DOMAIN      179    488       ROD.
FT   DOMAIN      489    643       TAIL.
FT   DOMAIN      179    214       COIL 1A.
FT   DOMAIN      215    233       LINKER 1.
FT   DOMAIN      234    325       COIL 1B.
FT   DOMAIN      326    349       LINKER 12.
FT   DOMAIN      350    488       COIL 2.
FT   SITE        432    432       STUTTER.
FT   DOMAIN        1    150       GLY/PHE/SER-RICH.
FT   DOMAIN      501    640       GLY/SER-RICH.
FT   MOD_RES      65     65       PHOSPHORYLATION (BY SIMILARITY).
FT   VARIANT     154    154       V -> G (IN EHK).
FT                                /FTId=VAR_003853.
FT   VARIANT     160    160       L -> P (IN EHK).
FT                                /FTId=VAR_003854.
FT   VARIANT     185    185       S -> P (IN EHK).
FT                                /FTId=VAR_003855.
FT   VARIANT     187    187       N -> S (IN EHK).
FT                                /FTId=VAR_003856.
FT   VARIANT     192    192       S -> P (IN EHK).
FT                                /FTId=VAR_003857.
FT   VARIANT     311    311       I -> V.
FT                                /FTId=VAR_003858.
FT   VARIANT     329    329       I -> T.
FT                                /FTId=VAR_003859.
FT   VARIANT     357    357       N -> Y.
FT                                /FTId=VAR_003860.
FT   VARIANT     489    489       E -> Q (IN EHK).
FT                                /FTId=VAR_003861.
FT   VARIANT     536    536       G -> C.
FT                                /FTId=VAR_003862.
FT   VARIANT     559    565       MISSING (IN ALLELE 1B).
FT                                /FTId=VAR_003864.
FT   VARIANT     632    632       R -> K (IN dbSNP:14024).
FT                                /FTId=VAR_003863.
SQ   SEQUENCE   643 AA;  65886 MW;  DF945DC462257850 CRC64;

  Query Match             3.7%;  Score 145;  DB 1;  Length 643;
  Best Local Similarity   20.3%;  Pred. No. 0.26;
  Matches   82;  Conservative   73;  Mismatches  166;  Indels   82;  Gaps   16;

Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGG-------GGSYSASSSSS 59
              ||| |||     | |||||||  :       ||  |||  ||       ||    : : |
Db        100 GGGFGGGGFGGGGIGGGGFGGFGSGGGGFGGGGFGGGGYGGGYGPVCPPGGIQEVTINQS 159

Qy         60 AAAAAGAAVLP----VKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTL 115
                        : |    ||  : | ::: :  |    :|     |||  :| :  :     |
```

```
Db        160 LLQPLNVEIDPEIQKVKSREREQIKSLNNQFASFIDK----VRFLEQQNQV--LQTKWEL 213

Qy        116 TYMSHRNSRTNIKRKTFK--VDDMLSKVEKMKGEQES-HSLSAHLQLTFTGFFHKND--- 169
                ::|||:     |:  ::::  :|:::| :|    |   ::|     : :| :
Db        214 LQQVDTSTRTHNLEPYFESFINNLRRRVDQLKSDQSRLDSELKNMQDMVEDYRNKYEDEI 273

Qy        170 KPSPNSENEQNSVTLEV---LLVKVCHKKRKD-----------VSCPIRQVPTGKKQVP 214
              |:|||  ::  :|    : ||  : : |                : |: |   :
Db        274 NKRTNAENEFVTIKKDVDGAYMTKVDLQAKLDNLQQEIDFLTALYQAELSQMQTQISETN 333

Qy        215 LIPDLNQTKPGNFPSL----------AVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNG 264
              :|  ::  :  :  |:             :: |  : :  |   |     | |
Db        334 VILSMDNNRSLDLDSIIAEVKAQNEDIAQKSKAEAESLYQSKYEELQITAGRHG------ 387

Qy        265 MINGETNENIDVNE-ELPARRKRNREDGEKTFVA--QMTVFDKNRRLQ--LLD-----GE 314
               :   |  |:::|     :| |:  |  |  ::  | :: |  :| :  | |      :
Db        388 --DSVRNSKIEISELNRVIQRLRSEIDNVKKQISNLQQSISDAEQRGENALKDAKNKLND 445

Qy        315 YEVAMQEMEE---------------CPISKKRATWETILDGK 341
               | |:|: :|               :  : ||: |:|:|:
Db        446 LEDALQQAKEDLARLLRDYQELMNTKLALDLEIATYRTLLEGE 488


Search completed: September  2, 2003, 11:02:14
Job time : 20.556 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:04:21 ; Search time 58.6986 Seconds
(without alignments)
1722.049 Million cell updates/sec

Title: US-09-874-162A-5
Perfect score: 3885
Sequence: 1 MAPQKHGGGGGGGSGPSAGS.....KALETDSVSGVSKQSKKQKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 510680 seqs, 136781880 residues

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3885	100.0	739	10	US-09-874-162A-5	Sequence 5, Appli
2	3413	87.9	776	10	US-09-874-162A-8	Sequence 8, Appli
3	2043	52.6	388	10	US-09-764-864-931	Sequence 931, App
4	1497	38.5	289	10	US-09-764-864-911	Sequence 911, App
5	1453	37.4	292	10	US-09-764-864-1369	Sequence 1369, Ap
6	919	23.7	175	10	US-09-764-864-1355	Sequence 1355, Ap
7	219.5	5.6	813	15	US-10-231-778-2	Sequence 2, Appli
8	162.5	4.2	1390	15	US-10-097-340-35	Sequence 35, Appl
9	151	3.9	669	15	US-10-286-264-40	Sequence 40, Appl
10	151	3.9	669	16	US-10-278-536-12	Sequence 12, Appl
11	146.5	3.8	1639	15	US-10-087-464-10	Sequence 10, Appl
12	144.5	3.7	1569	14	US-10-108-605-303	Sequence 303, App
13	142.5	3.7	474	10	US-09-736-457-1812	Sequence 1812, Ap
14	142.5	3.7	474	10	US-09-902-941-1812	Sequence 1812, Ap
15	142.5	3.7	474	10	US-09-849-626-1812	Sequence 1812, Ap
16	142.5	3.7	474	15	US-10-017-754-1812	Sequence 1812, Ap
17	142.5	3.7	474	15	US-10-205-823-381	Sequence 381, App
18	141	3.6	201	9	US-09-848-990-22	Sequence 22, Appl
19	141	3.6	201	10	US-09-760-364-14	Sequence 14, Appl
20	140.5	3.6	2845	12	US-10-140-472-207	Sequence 207, App
21	140.5	3.6	2845	12	US-10-141-761-207	Sequence 207, App
22	140.5	3.6	2845	12	US-10-142-885-207	Sequence 207, App
23	140.5	3.6	2845	15	US-10-123-155-207	Sequence 207, App
24	140.5	3.6	2845	16	US-10-146-731-207	Sequence 207, App
25	140	3.6	1968	12	US-10-140-472-163	Sequence 163, App
26	140	3.6	1968	12	US-10-141-761-163	Sequence 163, App
27	140	3.6	1968	12	US-10-142-885-163	Sequence 163, App
28	140	3.6	1968	15	US-10-123-155-163	Sequence 163, App
29	140	3.6	1968	16	US-10-146-731-163	Sequence 163, App
30	139.5	3.6	458	15	US-10-205-823-216	Sequence 216, App
31	138.5	3.6	357	9	US-09-864-761-35807	Sequence 35807, A
32	138.5	3.6	1670	12	US-10-140-472-325	Sequence 325, App
33	138.5	3.6	1670	12	US-10-141-761-325	Sequence 325, App
34	138.5	3.6	1670	12	US-10-142-885-325	Sequence 325, App
35	138.5	3.6	1670	15	US-10-123-155-325	Sequence 325, App
36	138.5	3.6	1670	16	US-10-146-731-325	Sequence 325, App
37	138	3.6	2388	15	US-10-184-644-457	Sequence 457, App
38	138	3.6	2388	15	US-10-184-634-457	Sequence 457, App
39	138	3.6	4185	12	US-10-140-472-67	Sequence 67, Appl
40	138	3.6	4185	12	US-10-141-761-67	Sequence 67, Appl
41	138	3.6	4185	12	US-10-142-885-67	Sequence 67, Appl
42	138	3.6	4185	15	US-10-123-155-67	Sequence 67, Appl
43	138	3.6	4185	16	US-10-146-731-67	Sequence 67, Appl
44	137	3.5	641	12	US-10-294-804-4	Sequence 4, Appli
45	137	3.5	641	16	US-10-138-098-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1
US-09-874-162A-5
; Sequence 5, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION: instant
; APPLICANT: Koontz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JJAZ1 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-5

Query Match 100.0%; Score 3885; DB 10; Length 739;
Best Local Similarity 100.0%; Pred. No. 6.6e-304;
Matches 739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60

Qy         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120

Qy        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180
```

```
Qy        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240

Qy        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300

Qy        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360

Qy        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420

Qy        421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480

Qy        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720

Qy        721 ALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||
Db        721 ALETDSVSGVSKQSKKQKL 739
```

RESULT 2
US-09-874-162A-8
; Sequence 8, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION:
; APPLICANT: Koontz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JJAZ1 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-8

instant

Query Match 87.9%; Score 3413; DB 10; Length 776;
Best Local Similarity 99.8%; Pred. No. 6.9e-266;
Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy         92 KPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESH 151
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 EPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESH 188

Qy        152 SLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKK 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 SLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKK 248

Qy        212 QVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETN 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 QVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETN 308

Qy        272 ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKR 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKR 368

Qy        332 ATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 ATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 428

Qy        392 GSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCP 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 GSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCP 488

Qy        452 WCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPG 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 WCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPG 548

Qy        512 FAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTC 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 FAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTC 608

Qy        572 LPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQ 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 LPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQ 668

Qy        632 MNHACMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 MNHACMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKG 728

Qy        692 ESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        729 ESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 776
```

RESULT 3
US-09-764-864-931
; Sequence 931, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 931
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-931

Query Match 52.6%; Score 2043; DB 10; Length 388;
Best Local Similarity 99.5%; Pred. No. 4.5e-156;
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        352 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 60

Qy        412 TRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCH 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCH 120

Qy        472 SRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVC 531
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 SRFIFNYVYHPKGARIDVSINEGYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVC 180
```

```
Qy        532 RPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEW 591
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        181 RPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHRDTCLPLRPQEMEVDSEDEKDPEW 240

Qy        592 LREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKN 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKN 300

Qy        652 LCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANG 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANG 360

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||||||||||||
Db        361 FSEINSKEKALETDSVSGVSKQSKKQKL 388


RESULT 4
US-09-764-864-911
; Sequence 911, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 911
;   LENGTH: 289
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-911

  Query Match             38.5%;  Score 1497;  DB 10;  Length 289;
  Best Local Similarity   99.6%;  Pred. No. 2.6e-112;
  Matches  283;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        456 NCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFS 515
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 DCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFS 65

Qy        516 RNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLR 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 RNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLR 125

Qy        576 PQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHA 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 PQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHA 185

Qy        636 CMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESAS 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESAS 245

Qy        696 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||||||||||||||||||||||||||||
Db        246 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 289


RESULT 5
US-09-764-864-1369
; Sequence 1369, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1369
;   LENGTH: 292
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (18)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (31)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (46)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (121)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (283)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1369

  Query Match             37.4%;  Score 1453;  DB 10;  Length 292;
  Best Local Similarity   96.8%;  Pred. No. 9.1e-109;
  Matches  275;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        196 RKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVT 255
              :|||||| |||||| | || ||||||||| |||||||||||||| |||||||||||||
Db          1 KKDVSCPIXQVPTGKXQXPLNPDLNQTKPGXFPSLAVSSNEFEPSXSHMVKSYSLLFRVT 60

Qy        256 RPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEY 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEY 120

Qy        316 EVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAK 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 XVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAK 180

Qy        376 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 240

Qy        436 NNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYV 479
              |||||||||||||||||||||||||||||||||||||||||||| |
Db        241 NNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNXV 284


RESULT 6
US-09-764-864-1355
; Sequence 1355, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1355
;   LENGTH: 175
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: SITE
;   LOCATION: (4)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (136)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (166)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1355

  Query Match             23.7%;  Score 919;  DB 10;  Length 175;
  Best Local Similarity   94.3%;  Pred. No. 4.1e-66;
  Matches  165;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        429 FYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARID 488
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FYQXLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARID 60

Qy        489 VSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESED 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESED 120

Qy        549 GEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEF 603
              ||||||||||||||| |||||||||||||||||||||||||| | : ::: |
Db        121 GEVEQQRTYSSGHNRXYFHSDTCLPLRPQEMEVDSEDEKDPEWLRXKPLHKLKSF 175


RESULT 7
US-10-231-778-2
; Sequence 2, Application US/10231778
; Publication No. US20030126647A1
; GENERAL INFORMATION:
;  APPLICANT: Bilodeau, Pierre
;  APPLICANT:  Chaudhury, Abdul M.
;  APPLICANT:  Dennis, Elizabeth S.
;  APPLICANT:  Koltunow, Anna M.G.
;  APPLICANT:  Luo, Ming
;  APPLICANT:  Peacock, William J.
;  TITLE OF INVENTION: Method for inducing seed development by down-regulating
;  TITLE OF INVENTION:  expression of the FIS2 gene
;  FILE REFERENCE: 72-98A
;  CURRENT APPLICATION NUMBER: US/10/231,778
;  CURRENT FILING DATE:  2002-11-08
;  PRIOR APPLICATION NUMBER: 09/398,237
;  PRIOR FILING DATE: 1999-09-20
;  PRIOR APPLICATION NUMBER: 60/101,184
;  PRIOR FILING DATE: 1998-09-21
;  PRIOR APPLICATION NUMBER: AU PP6061
;  PRIOR FILING DATE: 1998-09-22
;  PRIOR APPLICATION NUMBER: AU PP6062
;  PRIOR FILING DATE: 1998-09-22
;  PRIOR APPLICATION NUMBER: AU PP6063
;  PRIOR FILING DATE: 1998-09-22
;  PRIOR APPLICATION NUMBER: AU PQ1345
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: AU PQ1346
;  PRIOR FILING DATE: 1999-07-01
;  NUMBER OF SEQ ID NOS: 239
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 813
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-231-778-2

  Query Match              5.6%;  Score 219.5;  DB 15;  Length 813;
  Best Local Similarity   18.5%;  Pred. No. 7.6e-09;
  Matches  150;  Conservative  115;  Mismatches  306;  Indels  239;  Gaps   29;

Qy         77 EHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYM--SHRNSRTNIKRK-TFK 133
              : |  |  :    : ||  :|: | ||:   | || |:| |:  : :  ::|   |  |
Db         57 DDVDVDENII--KYIKPVAVYKKLETRSKNNPYFLRRSLKYIIQAKKKKKSNSGGKIRFN 114

Qy        134 VDDMLSKVEKMKGEQESHS------------LSAHL---------------------- 157
               |: :|:       |: |              |  ||
Db        115 YRDVSNKMTLKAEVVENFSCPFCLIPCGGHEGLQLHLKSSHDAFKFEFYRAEKDHGPEVD 174

Qy        158 -------------------QLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKV------ 191
                                ||:   |  ||       ::  |   | |||:::
Db        175 VSVKSDTIKFGVLKDDVGNPQLSPLTFCSKNRNQRRQRDDSNNVKKLNVLLMELDLDDDLP 234

Qy        192 -------CHKKRKDVSCPIR-------------QVPTGKKQVPLIPDLNQTKPGNF--P 228
                    |    :|| | |                |:   :   | :| :|    ||   |
Db        235 RGTENDSTHVNDDNVSSPPRAHSSEKISDILTTTQLAIAESSEPKVPHVND---GNVSSP 291

Qy        229 SLAVSSNE------------FEPSNSHMVKSYSLLF----RVTRPGRREFNGMINGETN 271
               | || |               |  :| ::           :: | :       : : |
Db        292 PRAHSSAEKNESTHVNDDDDVSSPPRAHSLEKNESTHVNEDNISSPPKAH-----SSKKN 346

Qy        272 ENIDVNEE---LPARRKRNREDGEKTFVAQMTVFDK--------NRRLQLLDGEYEVAMQ 320
              |:  :|:|    | | : ::|  :     |  : :          :|| ||    |
Db        347 ESTHMNDEDVSFPPRTRSSKETSDILTTTQPAIVEPSEPKVRRGSRRKQLYAKRY----- 401

Qy        321 EMEECPISKKRATWETILDGKRLPPF----ETFSQGP------------TLQFTLRWTG 363
                  | | |   | :          |  |  |                | |  :  :
Db        402 --------KARETQPAIAESSEPKVLHVNDENVSSPPEAHSLEKASDILTTTQPAIAESS 453

Qy        364 E-----TNDKSTAPIAKPLATRNSESLHQ--ENKPGSVKPTQTIAVKESLTTDLQTRKEK 416
              |      ||:: :   :  ::: ::|  :  :| |   |   :    : |||   |  |
Db        454 EPKVPHVNDENVSSTPRAHSSKKNKSTRKNVDNVPSPPKTRSSKKTSDILTTTQPTIAES 513

Qy        417 DTP------NENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLC 470
               |      ::|       :   | :||:     |::  |  | : :|  ::|   :
Db        514 SEPKVRHVNDDNVSSTPRAHSSKKNKSTRKND---DNIPSPPKTRSSKKTSNILTRTQ-- 568

Qy        471 HSRFIFNYVYHPKGARIDVSINECYDGSYAGNP--------QDIHRQPGFA--------- 513
                        |  |  :  :    |  :  |          :  |::   |
Db        569 -----------PAIAESEPKVPHVNDDKVSSTPRAHSSKKNKSTHKKDDNASLPPKTRSS 617

Qy        514 ------FSRNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFH 567
                   :   | |  |        || |: |  : | :|  | ::          | ::|
Db        618 KKTSDILATTQPAKAEP------SEPKVTRVSRRKELHAERCEAKRLERLK---GRQFYH 668

Qy        568 SDTCLPLRPQEM--EVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHG 625
              | |  |:  :::    |||:| |   |       ::|    |:: ||  | |||: | |
Db        669 SQTMQPMTFEQVMSNEDSENETDDYALDISERLRLERLVGVSKEEKRYMYLWNIFVRKQR 728

Qy        626 FIADNQMNHACMLFVE------------NYGQKIIKKNLCRNFMLHLVSMHDFNLISIMS 673
               |||  :  ||  | :            ::  :: :  |  | ::   : |    | : :
Db        729 VIADGHVPWACEEFAKLHKEEMKNSSSFDWWWRMFRIKLWNNGLICAKTFHKCTTILLSN 788

Qy        674 IDKAVTKLREMQQKLEKGESASPANEEITE 703
               |:|         :   | :|:  |::  |
Db        789 SDEA--------GQFTSGSAANANNQQSME 810


RESULT 8
US-10-097-340-35
; Sequence 35, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
;  APPLICANT: John MONAHAN
;  APPLICANT:  Manjula GANNAVARAPU
;  APPLICANT:  Sebastian HOERSCH
;  APPLICANT:  Shubhangi KAMATKAR
;  APPLICANT:  Steve G. KOVATS
;  APPLICANT:  Rachel E. MEYERS
;  APPLICANT:  Michael MORRISEY
;  APPLICANT:  Peter OLANDT
;  APPLICANT:  Ami SEN
;  APPLICANT:  Peter VEIBY
;  APPLICANT:  Gordon B. MILLS
;  APPLICANT:  Robert C. BAST, Jr.
;  APPLICANT:  Karen LU
;  APPLICANT:  Rosemarie SCHMANDT
```

```
;  APPLICANT:  Xumei ZHAO
;  APPLICANT:  Karen GLATT
;  TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
;  TITLE OF INVENTION:  Assessment, Prevention, and Therapy of Ovarian Cancer
;  FILE REFERENCE: MRI-030
;  CURRENT APPLICATION NUMBER: US/10/097,340
;  CURRENT FILING DATE:  2002-03-14
;  PRIOR APPLICATION NUMBER: 60/276,025
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/325,149
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/276,026
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/324,967
;  PRIOR FILING DATE: 2001/09/26
;  PRIOR APPLICATION NUMBER: 60/311,732
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: 60/325,102
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/323,580
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 363
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1390
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-097-340-35

  Query Match             4.2%;  Score 162.5;  DB 15;  Length 1390;
  Best Local Similarity   19.6%;  Pred. No. 0.00064;
  Matches  160;  Conservative  129;  Mismatches  282;  Indels  247;  Gaps   43;

Qy          8 GGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGG---GSCGGGGSYSASSSSSAAAAA 64
              | ||||:|  || | || ||| |     :||| :||    |  |||  |: | |
Db         84 GLGGGGAGAGAGVGAGGGGGSGA-----SSGGGAGGLQPSSRAGGGRPSSPSPS------ 132

Qy         65 GAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSR 124
                 |   :| ::| :| : |      :|  |:|           :|: | : |  :
Db        133 --VVSEKEKEELERLQKEEE----ERKKRLQLY-----------VFVMRCIAYPFNAKQP 175

Qy        125 TNIKRKTFKVDDMLSKVEK------MKGEQESHSLSAHLQL--TFTGFFHKNDKPSPN-- 174
              |:: |:  |:     :  |      : || :  :  | :    ::   | |:|: :
Db        176 TDMARRQQKISKQQLQTVKDRFQAFLNGETQIMADEAFMNAVQSYYEVFLKSDRVARMVQ 235

Qy        175 ----SENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSL 230
                  | |:   |             :| :          :|:|  :|:::        |
Db        236 SGGCSANDSREVF-----------KKHI----------EKRVRSLPEIDGLSKETVLSS 273

Qy        231 ------AVSSNEFEP----------SNSHMVKSYSLLFRVTRP--GRREFNGMI--NGET 270
                   |:   | :|          : | :: |   |: : :   | ::|   :  |
Db        274 WMAKFDAIYRGEEDPRKQQARMTASAASELILSKEQLYEMFQNILGIKKFEHQLLYNACQ 333

Qy        271 NENID-----VNEELPAR-RKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEE 324
               :| |      :  ||  | :  ::   |: |   ::   :|  :: |     : |  :|
Db        334 LDNPDEQAAQIRRELDGRLQMADQIARERKFPKFVSKEMENMYIEELKSSVNLLMANLES 393

Qy        325 CPISK---------KRATWETILDGKRLPPFETFSQGPTLQFTLRWT--GETNDKSTAP- 372
               |:||         ||:   :|:|       :       | |:|          || ||
Db        394 MPVSKGGEFKLQKLKRSHNASIIDMGEESENQLSKSDVVLSFSLEVVIMEVQGLKSLAPN 453

Qy        373 --IAKPLATRNSESLH----QENKP-----GSVKPTQTI-AVKESL---TTDLQTRKEKD 417
                 :   :     | |      : :||     |    |  : |||  |   :| :   ::|:
Db        454 RIVYCTMEVEGGEKLQTDQAEASKPTWGTQGDFSTTHALPAVKVKLFTESTGVLALEDKE 513

Qy        418 --------TPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKL 469
                     |||  :|          :   : | ::   :||   |  :     : |   :
Db        514 LGRVILHPTPNSPKQ-----------SEWHKMTVSK---NCPDQDLKIKLAVRMDKPQNM 559

Qy        470 CHSRFIF---NYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPIT 526
               || :::     |:     |  | :                |  ||
Db        560 KHSGYLWAIGKNVWKRWKKRFFVLVQV-------------SQYTFA------------- 592

Qy        527 HILVCRPKRTKASMSEFLESEDGEV---EQQRTYSSGHNRLYFHS----DTCLPLRPQEM 579
                 :|  :  ||   | |: :   |   : |     |  | :|::    || :
Db        593 ---MCSYREKKAEPQELLQLDGYTVDYTDPQPGLEGG--RAFFNAVKEGDTVI------- 640

Qy        580 EVDSEDEKDP-EWLR---------EKTI--TQIEEFS----DVNEGEKEVMKLWNLHVMK 623
                 |:||:|   |::          | :  ||::: :    :| : :  : : :     |
Db        641 -FASDDEQDRILWVQAMYRATGQSHKPVPPTQVQKLNAKGGNVPQLDAPISQFYADRAQK 699

Qy        624 HG---FIADNQMN--HACMLFV----------ENY--------GQKIIKKNLC-RNFML 658
              ||   ||: |  |  || : :            ::|        ||  :    | ||
Db        700 HGMDEFISSNPCNFDHASLFEMVQRLTLDHRLNDSYSCLGWFSPGQVFVLDEYCARN--- 756

Qy        659 HLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASP 696
                          :    : :  ||:: :: | |    |
Db        757 ----------GVRGCHRHLCYLRDLLERAENGAMIDP 783


RESULT 9
US-10-286-264-40
; Sequence 40, Application US/10286264
; Publication No. US20030093837A1
; GENERAL INFORMATION:
;  APPLICANT: Keddie, James
;  APPLICANT:  Riechmann, Jose-Luis
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Zhang, James
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Heard, Jacqueline
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Adam, Luc
;  APPLICANT:  Broun, Pierre
;  APPLICANT:  Reuber, Lynne
;  APPLICANT:  Pilgrim, Marsha
;  APPLICANT:  Samaha, Raymond
;  TITLE OF INVENTION: POLYNUCLEOTIDES FOR SEED TRAIT ALTERATION
;  FILE REFERENCE: MBI-008
;  CURRENT APPLICATION NUMBER: US/10/286,264
;  CURRENT FILING DATE:  2002-11-01
;  PRIOR APPLICATION NUMBER: 60/125,814
;  PRIOR FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 165
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
;   LENGTH: 669
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G634
US-10-286-264-40

  Query Match             3.9%;  Score 151;  DB 15;  Length 669;
  Best Local Similarity   19.0%;  Pred. No. 0.0019;
  Matches  145;  Conservative   94;  Mismatches  228;  Indels  298;  Gaps   35;

Qy          5 KHGGGGGG------GSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSS 58
              : ||||||        |  |:              :||   ||  |||  ||||| |:|| :
Db          2 EQGGGGGGNEVVEEASPISSRPPANNLEELMRFSAAADDGGLGGGGGGGGGSASSSSGN 61

Qy         59 SAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLH--RTLT 116
                    |:| ::                    : :    |   | ||:: |  | |
Db         62 RWPREETLALLRIR--------------------SDMDSTFRDATLKAPLWEHVSRKLL 100

Qy        117 YMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHK----NDKPS 172
              : :: |    | |   |      : :: :| :  |   |:       || :    | |
Db        101 ELGYKRSSKKCKEKFENVQKYYKRTKETRGGR--HDGKAY------KFFSQLEALNTTP- 151

Qy        173 PNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAV 232
              |:|       :|:|  : | :         ||    :      |:    :| :|
Db        152 PSS-------SLDVTPLSVAN--------PILMPSSSSSPFPV---FSQPQP-------- 185

Qy        233 SSNEFEPSNSHMVKSYSLLFRVTRP------------------GRREFNGMINGETNEN 273
                  : :|  :| |      |  | |                          :|| : : :::
```

```
Db        186 -QTQTQPPQTHNVS-----FTPTPPPLPLPSMGPIFTGVTFSSHSSSTASGMGSDDDDDD 239

Qy        274 IDVNEELPA------RRKRNREDGEKTFVAQMTVFD--------------------KNR 306
              :||::   |      |:: ||  | |     | :|:                     : |
Db        240 MDVDQANIAGSSSRKRKRGNRGGGGK----MMELFEGLVRQVMQKQAAMQRSFLEALEKR 295

Qy        307 RLQLLDGEYEVAMQEM-----EECPISKKRA----------------TWETILDGKRLPP 345
              : || |     |||       |   :|::||                |  ||    :|||
Db        296 EQERLDREEAWKRQEMARLAREHEVMSQERAASASRDAAIISLIQKITGHTI----QLPP 351

Qy        346 FETFSQGPTLQFTLRWTGETNDKSTAPIAK-PLATRNSES-------------------- 384
                :    |  |           |  :|: ||:|  |:|
Db        352 SLSSQPPPPYQ--------PPPAVTKRVAEPPLSTAQSQSQQPIMAIPQQQILPPPPPSH 403

Qy        385 --LHQENKPGSVKPTQTIAVKE--------------------SLTTDLQTRKEKDTP--- 419
                 ||  :    :| | : :                      :| : :: | : : |
Db        404 PHAHQPEQKQQQQPQQEMVMSSEQSSLPSSSRWPKAEILALINLRSGMEPRYQDNVPKGL 463

Qy        420 --NENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKH--------LKL 469
                 |    ::   :  || | ::  |         |  :|  | |  :|          |
Db        464 LWEEISTSMK---RMGYNRNAKRCKE-------KWENIN--KYYKKVKESNKKRPQDAKT 511

Qy        470 CHSRFIFNYVYHPKGARIDVSINECYDGSYAGN-----PQDIHRQPGFAFSRNGPVKRTP 524
              |           |   |:|:       ||  |:     |||                |::|
Db        512 C-----------PYFHRLDLLYRNKVLGSGGGSSTSGLPQD-------------QKQSP 546

Qy        525 ITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSE 584
              :|     :|             ::| |  |:|: |                  : :|   :
Db        547 VT---AMKP------------PQEGLVNVQQTHGSASTE----------EEEPIEESPQ 580

Qy        585 DEKDPEWLREKTITQ-----------------IEEFSDVNEGEKE 612
                : || |  : : |                 |||  : |  |:|
Db        581 GTEKPEDLVMRELIQQQQQLQQQESMIGEYEKIEESHNYNNMEEE 625


RESULT 10
US-10-278-536-12
; Sequence 12, Application US/10278536
; Publication No. US20030131386A1
; GENERAL INFORMATION:
;  APPLICANT: Samaha, Raymond
;  APPLICANT:  Heard, Jacqueline
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Reuber, Lynne
;  APPLICANT:  Riechmann, Jose-Luis
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Keddie, James
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Pilgrim, Marsha
;  APPLICANT:  Adam, Luc
;  APPLICANT:  Broun, Pierre
;  TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
;  FILE REFERENCE: MBI-011
;  CURRENT APPLICATION NUMBER: US/10/278,536
;  CURRENT FILING DATE:  2002-10-22
;  PRIOR APPLICATION NUMBER: 60/125,814
;  PRIOR FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 669
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G634
US-10-278-536-12

  Query Match           3.9%;  Score 151;  DB 16;  Length 669;
  Best Local Similarity   19.0%;  Pred. No. 0.0019;
  Matches  145;  Conservative   94;  Mismatches  228;  Indels  298;  Gaps   35;

Qy          5 KHGGGGGG------GSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSS 58
              : ||||||      |  |:             :||   ||  |||  ||||| |:|| :
Db          2 EQGGGGGGNEVVEEASPISSRPPANNLEELMRFSAAADDGGLGGGGGGGGGSASSSSGN 61

Qy         59 SAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLH--RTLT 116
                     |:| ::                    : :    |   | ||:: |  | |
Db         62 RWPREETLALLRIR--------------------SDMDSTFRDATLKAPLWEHVSRKLL 100

Qy        117 YMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHK----NDKPS 172
               : :: |    | |   |     : :: :| :  |   |:       || :    |  |
Db        101 ELGYKRSSKKCKEKFENVQKYYKRTKETRGGR--HDGKAY------KFFSQLEALNTTP- 151

Qy        173 PNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAV 232
              |:|       :|:|  : | :        ||    :     |:    :| :|
Db        152 PSS-------SLDVTPLSVAN--------PILMPSSSSSPFPV---FSQPQP-------- 185

Qy        233 SSNEFEPSNSHMVKSYSLLFRVTRP------------------GRREFNGMINGETNEN 273
                : :|  :| |      |  | |                          :|| : : :::
Db        186 -QTQTQPPQTHNVS-----FTPTPPPLPLPSMGPIFTGVTFSSHSSSTASGMGSDDDDDD 239

Qy        274 IDVNEELPA------RRKRNREDGEKTFVAQMTVFD--------------------KNR 306
              :||::   |      |:: ||  | |     | :|:                     : |
Db        240 MDVDQANIAGSSSRKRKRGNRGGGGK----MMELFEGLVRQVMQKQAAMQRSFLEALEKR 295

Qy        307 RLQLLDGEYEVAMQEM-----EECPISKKRA----------------TWETILDGKRLPP 345
              : || |     |||       |   :|::||                |  ||    :|||
Db        296 EQERLDREEAWKRQEMARLAREHEVMSQERAASASRDAAIISLIQKITGHTI----QLPP 351

Qy        346 FETFSQGPTLQFTLRWTGETNDKSTAPIAK-PLATRNSES-------------------- 384
                :    |  |           |  :|: ||:|  |:|
Db        352 SLSSQPPPPYQ--------PPPAVTKRVAEPPLSTAQSQSQQPIMAIPQQQILPPPPPSH 403

Qy        385 --LHQENKPGSVKPTQTIAVKE--------------------SLTTDLQTRKEKDTP--- 419
                 ||  :    :| | : :                      :| : :: | : : |
Db        404 PHAHQPEQKQQQQPQQEMVMSSEQSSLPSSSRWPKAEILALINLRSGMEPRYQDNVPKGL 463

Qy        420 --NENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKH--------LKL 469
                 |    ::   :  || | ::  |         |  :|  | |  :|          |
Db        464 LWEEISTSMK---RMGYNRNAKRCKE-------KWENIN--KYYKKVKESNKKRPQDAKT 511

Qy        470 CHSRFIFNYVYHPKGARIDVSINECYDGSYAGN-----PQDIHRQPGFAFSRNGPVKRTP 524
              |           |   |:|:       ||  |:     |||                |::|
Db        512 C-----------PYFHRLDLLYRNKVLGSGGGSSTSGLPQD-------------QKQSP 546

Qy        525 ITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSE 584
              :|     :|             ::| |  |:|: |                  : :|   :
Db        547 VT---AMKP------------PQEGLVNVQQTHGSASTE----------EEEPIEESPQ 580

Qy        585 DEKDPEWLREKTITQ-----------------IEEFSDVNEGEKE 612
                : || |  : : |                 |||  : |  |:|
Db        581 GTEKPEDLVMRELIQQQQQLQQQESMIGEYEKIEESHNYNNMEEE 625


RESULT 11
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
;  APPLICANT: Chishti, Athar
;  APPLICANT:  Oh, Steven
;  APPLICANT:  Liu, David
;  APPLICANT:  Goel, Vikas
;  APPLICANT:  Li, Xuerong
;  TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
;  FILE REFERENCE: S1237/7019
;  CURRENT APPLICATION NUMBER: US/10/087,464
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 06/272,930
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 1639
;   TYPE: PRT
```

; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match 3.8%; Score 146.5; DB 15; Length 1639;
Best Local Similarity 20.6%; Pred. No. 0.016;
Matches 168; Conservative 123; Mismatches 291; Indels 235; Gaps 46;

```
Qy         10 GGGGSGPSAGSGGG-GFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAV 68
Db         68 GSKGSVASGGSGGSVASGGSVASGGSVASGGSVASGGSGNSRRTNPSDNSSDSDAKSYA- 126

Qy         69 LPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIK 128
Db        127 ------DLKHRVRNYLLTIKELKYP-QLFDL--TNHML-------TLC--------DNIH 162

Qy        129 RKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLL 188
Db        163 GFKYLID----------GYEEINELLYKLNFYFD--------------------LLRAKL 192

Qy        189 VKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSY 248
Db        193 NDVC----ANDYC----------QIPF-------------NLKIRANELD-----VLK-- 218

Qy        249 SLLFRVTRP--------GRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
Db        219 KLVFGYRKPLDNIKDNVGKMEDYIKKNKKTIENI--NELI--------EESKKTIDKNKN 268

Qy        301 VFDKNRRLQLLDGEYEVAM--QEMEEC----------PISKKRATWETILDG----KRL 343
Db        269 ATKEEEKKKLYQAQYDLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNP 328

Qy        344 PPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSES----------LHQENKPG 392
Db        329 PPANS-GNTPNTLLDKNKKIEEHEKEIKEIAKTIKF-NIDSLFTDPLELEYYLREKNK-- 384

Qy        393 SVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN--NNTRQQTEARDDLHC 450
Db        385 ------NIDISAKVET-----KESTEPNEYPN--GVTYPLSYNDINNALNELNSFGDLIN 431

Qy        451 PW----------CTLNCRKLY--SLLKHLKLCHSRFIFN-YVYHPKGARI-DVS------ 490
Db        432 PFDYTKEPSKNIYTDNERKKFINEIKEKIKIEKKKIESDKKSYEDRSKSLNDITKEYEKL 491

Qy        491 INECYDGSYAGNPQ--DIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEF---LE 545
Db        492 LNEIYDSKFNNNIDLTNFEKMMGKRYS----YKVEKLTH------HNTFASYENSKHNLE 541

Qy        546 SEDGEVEQQRTYSSGH----NRLYFHSDTCLPLRPQEMEVDSED-EKDPEWLREKTITQI 600
Db        542 KLTKALKYMEDYSLRNIVVEKELKYYKNLISKIE-NEIETLVENIKKDEEQLFEKKITK- 599

Qy        601 EEFSDVNEGEKEVMKLWNL-HVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLH 659
Db        600 ----DENKPDEKILEVSDIVKVQVQKVLLMNKID-------ELKKTQLILKNVELKHNIH 648

Qy        660 LVSMH-------DFNLISIMSIDKAVTKLREMQQKLE----------KGESASPANEEIT 702
Db        649 VPNSYKQENKQEPYYLI---VLKKEIDKLKVFMPKVESLINEEKKNIKTEGQSDNSEPST 705

Qy        703 EEQNGTANGFSEINSKEK---ALETDSVSGVSKQSKK 736
Db        706 E---GEITGQATTKPGQQAGSALEGDSVQAQAQEQKQ 739
```

RESULT 12
US-10-108-605-303
; Sequence 303, Application US/10108605
; Publication No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD
; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 303
; LENGTH: 1569
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-303

Query Match 3.7%; Score 144.5; DB 14; Length 1569;
Best Local Similarity 19.5%; Pred. No. 0.022;
Matches 104; Conservative 69; Mismatches 188; Indels 173; Gaps 22;

```
Qy           9 GGGGGSGPSAGSGGGG----------------FGGSAAVAAATASGGKSGGGSCGGGGSY 52
Db         849 GGVGGQGGSRRSGDAAKRRYIRRKERNNNSNKEAGSGAKAAAVAAGGAGGNGSSSVGGGD 908

Qy          53 SASSSSSAAAAAGAAVLPVKKPKMEHV-----QADHELFLQAFEKPTQIYRFLRTRNLIA 107
Db         909 SGNGTGTATTS--SSTTPTTTPLVANTARGKRQPQHP---TGANSSSSLWPDSRQRQ--- 960

Qy         108 PIFLHRTLTYMSHRNSRTNIK-----RKTFKVDDMLSKVEKMKGEQESHSL---SAHLQL 159
Db         961 --------RQRHHSPSSTSISALSGAKKARKSATNSSSSNSTQQQQNHHGTKINNHHL-- 1010

Qy         160 TFTGF---------FHKNDKPSPNSENEQNS---------VTLEVLLVKVCHKKRKDVSC 201
Db        1011 --NGYGDQWGEQSRSRRNSSPSNSHKNERTSERRVRPIYDINNIVIPYSMLAQSKMEI-I 1067

Qy         202 PIRQVP-----------------------------------TGKKQVPLIPDLNQTKP 224
Db        1068 PYKEIPIPKWRIVDSDNDKGKHSSDESMERKLSNGCVASKSEELAKQQLPPPPDQDQEQP 1127

Qy         225 GNFP--------SLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDV 276
Db        1128 ALQPPKVNGLKENQAVKHNNNNNTNNNNNK---------------NGLVNGNAKK---- 1167

Qy         277 NEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWET 336
Db        1168 -DELKAAEDHDTQSKEDDDPKEKVVKPK------LNG-------NLIKNPIDKTAETQDK 1213

Qy         337 ILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAP---IAKPLATRNSESLHQENKPGS 393
Db        1214 IAQPAPEPP---LPKRPKLE-----TPTSNEVTKANGQLVQQQLESRESEEEPQH----- 1260

Qy         394 VKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDD 447
Db        1261 --------VDEDLSDEAFIMRHQRALLEERRRFETFLKFPW--STRSRANRRVD 1304
```

RESULT 13
US-09-736-457-1812
; Sequence 1812, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun

```
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1812
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-736-457-1812

  Query Match             3.7%;  Score 142.5;  DB 10;  Length 474;
  Best Local Similarity   39.0%;  Pred. No. 0.0055;
  Matches   41;  Conservative   10;  Mismatches   25;  Indels   29;  Gaps    5;

Qy          6 HGGGGGGGSGPSAGSGGGGFGGSA----------AVAAATASGGKS---------GGGS 45
              |||||||||  : | |||  || |           :  |  | || |         |||
Db        166 HGGGGGGGSSNAGGGGGGASGGGANSKPAQKKSCGSKVAGGAGGGVSKPHAKLILAGGG- 224

Qy         46 CGGGGSYSASSSSSAAAAAG-AAVLPVKKPKMEHVQADHELFLQA 89
                ||| : :|:::| ||  || ||:||:         |||    :|
Db        225 -GGGKAAAAAAASFAAEQAGAAALLPL------GAAADHHSLYKA 262


RESULT 14
US-09-902-941-1812
; Sequence 1812, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C17
;  CURRENT APPLICATION NUMBER: US/09/902,941
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 2002
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1812
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-902-941-1812

  Query Match             3.7%;  Score 142.5;  DB 10;  Length 474;
  Best Local Similarity   39.0%;  Pred. No. 0.0055;
  Matches   41;  Conservative   10;  Mismatches   25;  Indels   29;  Gaps    5;

Qy          6 HGGGGGGGSGPSAGSGGGGFGGSA----------AVAAATASGGKS---------GGGS 45
              |||||||||  : | |||  || |           :  |  | || |         |||
Db        166 HGGGGGGGSSNAGGGGGGASGGGANSKPAQKKSCGSKVAGGAGGGVSKPHAKLILAGGG- 224

Qy         46 CGGGGSYSASSSSSAAAAAG-AAVLPVKKPKMEHVQADHELFLQA 89
                ||| : :|:::| ||  || ||:||:         |||    :|
Db        225 -GGGKAAAAAAASFAAEQAGAAALLPL------GAAADHHSLYKA 262


RESULT 15
US-09-849-626-1812
; Sequence 1812, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
;  APPLICANT: Bangur, Chaitanya
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Switzer, Anne
;  APPLICANT:  McNeill, Patricia
;  APPLICANT:  Clapper, Jonathan
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C16
;  CURRENT APPLICATION NUMBER: US/09/849,626
;  CURRENT FILING DATE:  2001-05-03
;  NUMBER OF SEQ ID NOS: 1926
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1812
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-849-626-1812

  Query Match             3.7%;  Score 142.5;  DB 10;  Length 474;
  Best Local Similarity   39.0%;  Pred. No. 0.0055;
  Matches   41;  Conservative   10;  Mismatches   25;  Indels   29;  Gaps    5;

Qy          6 HGGGGGGGSGPSAGSGGGGFGGSA----------AVAAATASGGKS---------GGGS 45
              |||||||||  : | |||  || |           :  |  | || |         |||
Db        166 HGGGGGGGSSNAGGGGGGASGGGANSKPAQKKSCGSKVAGGAGGGVSKPHAKLILAGGG- 224

Qy         46 CGGGGSYSASSSSSAAAAAG-AAVLPVKKPKMEHVQADHELFLQA 89
                ||| : :|:::| ||  || ||:||:         |||    :|
Db        225 -GGGKAAAAAAASFAAEQAGAAALLPL------GAAADHHSLYKA 262


Search completed: September  2, 2003, 11:18:12
Job time : 61.6986 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37 ; Search time 24.834 Seconds
(without alignments)
2861.747 Million cell updates/sec

Title: US-09-874-162A-5
Perfect score: 3885
Sequence: 1 MAPQKHGGGGGGGSGPSAGS.....KALETDSVSGVSKQSKKQKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	168.5	4.3	3968	2	A44265	trithorax homolog
2	157	4.0	622	2	I37984	keratin 9, type I,
3	154.5	4.0	1289	2	I84505	calcium-dependent
4	153.5	4.0	1711	1	A47392	chromodomain-helic
5	153	3.9	549	2	B32372	male-specific doub
6	153	3.9	581	1	KRMS2	keratin, type II c
7	152	3.9	427	2	A32372	female-specific do
8	151.5	3.9	393	2	T20268	hypothetical prote
9	151	3.9	464	2	A56600	intermediate filam
10	149.5	3.8	2639	2	T31328	fibroin - Chinese
11	148.5	3.8	643	1	KRHU2	keratin 1, type II
12	148.5	3.8	1611	2	T06677	hypothetical prote
13	148	3.8	863	2	T51002	hypothetical prote
14	146.5	3.8	1235	1	PWBYH	potassium transpor
15	146.5	3.8	1639	2	S05603	major merozoite su
16	145	3.7	1744	2	JH0720	tanabin - African
17	144.5	3.7	1640	2	A24594	probable major sur
18	144.5	3.7	1875	2	S38173	myosin-like protei
19	144	3.7	592	2	E82759	endo-1,4-beta-gluc
20	144	3.7	594	2	B86456	protein trihelix D
21	143.5	3.7	420	2	A37343	keratin 13, type I
22	143	3.7	270	2	A60830	keratin, 70k type
23	143	3.7	385	2	T20410	hypothetical prote
24	142.5	3.7	474	2	I38240	transcription fact
25	142.5	3.7	1226	2	T24045	hypothetical prote
26	142.5	3.7	2038	2	A43742	female sterile hom
27	142	3.7	1119	2	B70126	surface-located me
28	141.5	3.6	255	2	B84777	hypothetical prote
29	141.5	3.6	291	1	S31415	glycine-rich prote
30	141.5	3.6	582	2	E84771	fertilization-inde
31	141.5	3.6	1087	2	T30844	serine-repeat anti
32	140	3.6	338	1	KNMU	glycine-rich cell
33	140	3.6	569	1	KRMSE1	keratin, 59K type
34	140	3.6	570	2	S07330	keratin, epidermal
35	139.5	3.6	912	2	B44450	ubiquitin-specific
36	139.5	3.6	1085	2	S62516	hypothetical coile
37	139	3.6	171	2	H84709	probable glycine-r
38	139	3.6	252	1	S01821	glycine-rich prote
39	139	3.6	419	2	A25438	keratin, type I cy
40	139	3.6	481	2	A35628	loricrin - mouse
41	139	3.6	1585	2	T31611	hypothetical prote
42	139	3.6	2500	1	WMHUE2	HIV-EP2 enhancer-b
43	138	3.6	1701	2	A26868	major merozoite su
44	138	3.6	5170	2	T15348	hypothetical prote
45	137.5	3.5	396	2	T49109	glycine-rich prote

ALIGNMENTS

RESULT 1
A44265
trithorax homolog HTX, version 2 - human
N;Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)
C;Species: Homo sapiens (man)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000
C;Accession: A44265; A44264; I58112; I37165; I38485
R;Tkachuk, D.C.; Kohler, S.; Cleary, M.L.
Cell 71, 691-700, 1992
A;Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal transl
A;Reference number: A44265; MUID:93046667; PMID:1423624
A;Accession: A44265
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3968 <TKA>
A;Cross-references: GB:L04284; NID:g184393; PIDN:AAA58669.1; PID:g184394
A;Note: sequence extracted from NCBI backbone (NCBIP:117729)
R;Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Canaani, O.; Cimino, G.; Croce, C.M.;
Cell 71, 701-708, 1992
A;Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-
A;Reference number: A44264; MUID:93046668; PMID:1423625
A;Accession: A44264
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 63-316,'GLLINSELEK',327,'Q',329,'VR',332,'DKEGTPP',340,'T',342,'EDKTVVRQS
546,'LQIESTSP',2555-3554,'N',3556-3594,'V',3596-3899,'A',3901-3968 <GU1>
A;Cross-references: GB:L04731; NID:g339921
A;Note: sequence extracted from NCBI backbone (NCBIP:117779)
R;Djabali, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.
Nature Genet. 2, 113-118, 1992
A;Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in a
A;Reference number: I58112; MUID:93265134; PMID:1303259
A;Accession: I58112
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1317-1700,'DD',1703-1936,'H',1938-2180,'S',2182-2328 <DJA>
A;Cross-references: GB:L01986; NID:g307522; PIDN:AAA92511.1; PID:g553800
R;Marschalek, R.; Greil, J.; Lochner, K.; Nilson, I.; Siegler, G.; Zweckbronner, I.;
Br. J. Haematol. 90, 308-320, 1995
A;Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in t
A;Reference number: I37165; MUID:95315013; PMID:7794749
A;Accession: I37165
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1212-1603,'GTE' <MAR>
A;Cross-references: EMBL:X83604; NID:g897757; PIDN:CAA58584.1; PID:g899268
A;Note: submitted to the EMBL/GenBank/DDBJ databases by R. Marschalek, 20 December 19
R;Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Canaani, O.; Saito, H
Cancer Res. 54, 2327-2330, 1994
A;Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involve
A;Reference number: I38485
A;Accession: I38485

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1251-1486,'G',1488-1538 <RES>
A;Cross-references: EMBL:U04737; NID:g451554; PID:g451555
C;Genetics:
A;Gene: GDB:MLL; HTX; ALL-1; HRX
A;Cross-references: GDB:128819; OMIM:159555
A;Map position: 11q23-11q23
A;Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
A;Note: the list of introns is incomplete
C;Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
F;1434-1456/Region: zinc finger CCHC motif
F;1479-1506/Region: zinc finger CCHC motif
F;1527-1556/Region: zinc finger CCCC motif
F;1569-1596/Region: zinc finger CCHC motif
F;1873-1900/Region: zinc finger CCHC motif
F;1933-1955/Region: zinc finger CCHC motif

  Query Match             4.3%;  Score 168.5;  DB 2;  Length 3968;
  Best Local Similarity   20.5%;  Pred. No. 0.098;
  Matches  180;  Conservative   97;  Mismatches  312;  Indels  287;  Gaps   39;

Qy          3 PQKHGGGGGGG-----------------SGPSAGSGGGGFGGS-AAVAAATASGGKSGG 43
              |   |||||||                 ||  | || |   |  ||||| |: | ||
Db         13 PGTTGGGGGGGRRGLGGAPRQRVPALLLPPGPPVGGGGPGAPPSPPAVAAAAAAAGSSGA 72

Qy         44 GSCGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTR 103
              |  ||  : ||:|||||:::: ::      | :  |    :  ||
Db         73 GVPGGAAAASAASSSSASSSSSSSSSASSGPALLRVGPGFDAALQ--------------- 117

Qy        104 NLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTG 163
               : | |               ||::|  |:     |      || |           | |
Db        118 -VSAAI-------------GTNLRR--FRAVFGESGGGGGSGEDEQ----------FLG 150

Qy        164 FFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTK 223
              |          |: |         |:|          : ||    |   |   : :
Db        151 F---------GSDEE----------VRV-------------RSPTRSPSVKTSPRKPRGR 178

Qy        224 P--GNFPSLAVSSNE--FEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEE 279
              |  |:  : |: |:   | | |    ||                          : |   :
Db        179 PRSGSDRNSAILSDPSVFSPLNKSETKS----------------------GDKIKKKDS 215

Qy        280 LPARRKRNREDGEKTFVAQMTVFDKNRRL-QLLDGEYEVAMQEMEECP---------ISK 329
                 :|| |     ||          : : :|  |  | ::::::  |         | |
Db        216 KSIEKKRGR---PPTFPGVKIKITHGKDISELPKGNKEDSLKKIKRTPSATFQQATKIKK 272

Qy        330 KRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESL---- 385
              ||      .||  |     |  |  ||   :       :   |  : : | :  ||
Db        273 LRA-------GKLSPLKSKFKTG-KLQIGRKGVQIVRRRGRPPSTERIKTPSVSSLILNW 324

Qy        386 HQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIF------YQFLYNNNTR 439
               | |  :       || |    : |:  |   |   :| :|         | | :|: |
Db        325 KSPRKSGKTR-------KEHLHLQKKIRQLSDKALEGLSQLGLFLLQKGQMQPLLSNSYR 377

Qy        440 -QQTEARDDLHCPWCTLNCRKLYSLLKHLK--------------LCHSRFIFNYVYHP- 482
              |:  |:  :       |  ||: : :|::                    |||  :  | |
Db        378 GQKKGAQKKIEKEAAQLQGRKVKTQVKNIRQFIMPVVSAISSRIIKTPRRFIEDEDYDPP 437

Qy        483 -KGARIDVSINECYDGSYAGNPQ--------------DIHR--QPGFAFSRNG------- 518
              | ||:: : |  :     |: :                |  |    |    | :
Db        438 IKIARLESTPNSRFSAPSCGSSEKSSAASQHSSQMSSDSSRSSSPSVDTSTDSQASEEIQ 497

Qy        519 ----------------PVKRTPITHILVCRPKRTKASMSEF----------LESEDGEVE 552
                             |: ::|       | :|   |   |          |:|
Db        498 VLPEERSSTPEVHPPLPISQSPENESNDRRSRRYSVSERSFGSRTTKKLSTLQS----AP 553

Qy        553 QQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEF-----SDVN 607
              ||:| ||    |        ||:|      |  :  | ||   ||     |     : :
Db        554 QQQTSSSPPPPLLTPPP---PLQP----ASSISDHTP-WLMPPTIPLASPFLPASTAPMQ 605

Qy        608 EGEKEVMK----LWNLHVMKHGFIADNQMNHACMLFVENYGQK-IIKKNLCRNFMLHLVS 662
                | :::      |    :||      : |       | :: :|:|. :  ||    ::
Db        606 GKRKSILREPTFRWT--SLKHSRSEPQYFSSA------KYAKEGLIRKPIFDNFRPPPLT 657

Qy        663 MHDFNLISIMSIDKAVTKLR----------EMQQ--------KLEKGESASPANEEITE 703
                |    |  |          |          :| :         :    |: |   | :|
Db        658 PEDVGFASGFSASGTAASARLFSPLHSGTRFDMHKRSPLLRAPRFTPSEAHSRIFESVTL 717

Qy        704 EQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
                | |: | |              |||| : :|:|:
Db        718 PSNRTSAGTSS------------SGVSNRKRKRKV 740


RESULT 2
I37984
keratin 9, type I, cytoskeletal - human
N;Alternate names: cytokeratin 9; scatter protein 60K chain, placental
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Oct-1999
C;Accession: I37984; S40307; S77921; S41161; B35494; I37943
R;Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Sch
Nature Genet. 6, 174-179, 1994
A;Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).
A;Reference number: I37984; MUID:94214498; PMID:7512862
A;Accession: I37984
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-622 <RES>
A;Cross-references: EMBL:X75015; NID:g453154; PIDN:CAA52924.1; PID:g453155
R;Langbein, L.; Heid, H.W.; Moll, I.; Franke, W.W.
Differentiation 55, 57-71, 1993
A;Title: Molecular characterization of the body site-specific human epidermal cytoker
A;Reference number: I37943; MUID:94131202; PMID:7507869
A;Accession: S40307
A;Molecule type: mRNA
A;Residues: 1-622 <LAN>
A;Cross-references: EMBL:Z29074; NID:g435475; PIDN:CAA82315.1; PID:g435476
A;Accession: S77921
A;Molecule type: protein
A;Residues: 29-53;62-104;167-188;199-233;241-249;295-362;374-430;450-480;579-604 <LAF
R;Langbein, L.
submitted to the EMBL Data Library, December 1993
A;Reference number: S41161
A;Accession: S41161
A;Molecule type: mRNA
A;Residues: 1-11,'SR',13-622 <LAW>
A;Cross-references: EMBL:Z29074; NID:g435475; PIDN:CAA82315.1; PID:g435476
R;Rosen, E.M.; Meromsky, L.; Romero, R.; Setter, E.; Goldberg, I.
Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990
A;Title: Human placenta contains an epithelial scatter protein.
A;Reference number: A35494; MUID:90267446; PMID:2140676
A;Accession: B35494
A;Molecule type: protein
A;Residues: 'X',450-465 <ROS>
C;Genetics:
A;Gene: GDB:KRT9; EPPK
A;Cross-references: GDB:303970; OMIM:144200
A;Map position: 17q12-17q21
A;Introns: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2
A;Note: defects in this gene may cause epidermolytic palmoplantar keratoderma
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil; intermediate filament
F;1-153/Domain: head #status predicted <HEA>
F;154-459/Domain: helical rod #status predicted <ROD>
F;460-622/Domain: tail #status predicted <TAI>

  Query Match             4.0%;  Score 157;  DB 2;  Length 622;
  Best Local Similarity   20.9%;  Pred. No. 0.042;
  Matches   89;  Conservative   63;  Mismatches  165;  Indels  108;  Gaps   19;

Qy          9 GGGGGSGPSAGSGGGGFGGSAAVAAATASGGKS---------GGGSCGG-GGSYSA---- 54
              ||| | | || | ||||||  :      : || |         |||| || || | :
Db         73 GGGSGGGFSASSLGGGFGGGSRGFGGASGGGYSSSGGFGGGFGGGSGGGFGGGYGSGFGG 132

Qy         55 --SSSSSAAAAAGAAVLPVKKPKMEHVQADHELFL---QAFEK---------------- 92
                    |    |  :   :|  |: : :    :|   || |:
```

```
Db        133 LGGFGGGAGGGDGGILTANEKSTMQELNSRLASYLDKVQALEEANNDLENKIQDWYDKKG 192

Qy         93 PTQIYR----FLRT----RNLIAPIFLHRTLTYMSHRNSRTNIK--RKTFKVDDMLSK-- 140
Db        193 PAAIQKNYSPYYNTIDDLKDQIVDLTVGNNKTLLDIDNTRMTLDDFRIKFEMEQNLRQGV 252

Qy        141 --------------VEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLE 185
Db        253 DADINGLRQVLDNLTMEKSDLEMQYETLQEELMA-----LKKNHKEEMSQLTGQNS---- 303

Qy        186 VLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMV 245
Db        304 -----------GDVNVEINVAP-GKDLTKTLNDMRQ----EYEQL-IAKNRKDIENQYET 346

Qy        246 KSYSLLFRVTRPGR------REFNGMINGETNENIDVNEEL--PARRKRNREDGEKTFVA 297
Db        347 QITQIEHEVSSSGQEVQSSAKEVTQLRHGVQELEIELQSQLSKKAALEKSLEDTKNRYCG 406

Qy        298 QMTVFDK---NRRLQLLDGEYEVAMQEME-------ECPISKKRATWETILDGKRLPPFE 347
Db        407 QLQMIQEQISNLEAQITDVRQEIECQNQEYSLLLSIKMRLEKEIETYHNLLEGGQ-EDFE 465

Qy        348 TFSQG 352
Db        466 SSGAG 470
```

RESULT 3
I84505
calcium-dependent actin-binding protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 02-Aug-2002
C;Accession: I84505
R;Walent, J.H.; Porter, B.W.; Martin, T.F.
Cell 70, 765-775, 1992
A;Title: A novel 145 kd brain cytosolic protein reconstitutes Ca(2+)-regulated secretion
A;Reference number: I39188; MUID:92386596; PMID:1516133
A;Accession: I84505
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1289 <RES>
A;Cross-references: EMBL:U16802; NID:g577427; PIDN:AAB88635.1; PID:g577428
C;Genetics:
A;Gene: CAPS
C;Superfamily: calcium-dependent actin-binding protein; pleckstrin repeat homology
C;Keywords: actin binding

  Query Match   4.0%; Score 154.5; DB 2; Length 1289;
  Best Local Similarity   19.1%; Pred. No. 0.16;
  Matches 157; Conservative 130; Mismatches 283; Indels 251; Gaps   43;

```
Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGG---GSCGGGGSYSASSSSSAAAA 63
Db         44 GSAGMGGSGAGAGVGAGGGGGSGA-----SSGGGAGGLQPSSRAGGGRPSSPSPS----- 93

Qy         64 AGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNS 123
Db         94 ---VVSEKEKEELERLQKEEE----ERKKRLQLY----------VFVMRCIAYPFNAKQ 135

Qy        124 RTNIKRKTFKVDDMLSKVEK------MKGEQESHSLSAHLQL--TFTGFFHKNDKPSPN- 174
Db        136 PTDMARRPRKISKQQLQTVKDRFQAFLNGETQIVADEAFMNAVQSYYEVFLKSDRVARMV 195

Qy        175 -----SENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPS 229
Db        196 QSGGCSANDSREVF-----------KKHI----------EKRVRSLPEIDGLSKETVLS 233

Qy        230 L------AVSSNEFEP----------SNSHMVKSYSLLFRVTRP--GRREFNGMI--NGE 269
Db        234 SWMAKFDAIYRGEEDPRKQQARMTASAASELILSKEQLYEMFQNILGIKKFEHQLLYNAC 293

Qy        270 TNENID-----VNEELPAR-RKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEME 323
Db        294 QLDNPDEQAAQIRRELDGRLQMADQIARERKFPKFVSKEMENMYIEELKSSVNLLMANLE 353

Qy        324 ECPISK---------KRATWETILDGKRLPPFETFSQGPTLQFTLRWT--GETNDKSTAP 372
Db        354 SMPVSKGGEFKLQKLKRSHNASIIDMGEESENQLSKSDVLLSFSLEVVIMEVQGLKSLAP 413

Qy        373 ---IAKPLATRNSESLH----QENKP-----GSVKPTQTI-AVKESL---TTDLQTRKEK 416
Db        414 NRIVYCTMEVEGGEKLQTDQAEASKPTWGTQGDFSTTHALPAVKVKLFTESTGVLALEDK 473

Qy        417 D--------TPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLK 468
Db        474 ELGRVILHPTPNSPKQ-----------SEWHKMTVSK---NCPDQDLKIKLAVRMDKPQN 519

Qy        469 LCHSRFIFN-----YVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRT 523
Db        520 MKHSGYLWTIGKNVWKRWKKRFFVLVQVSQ-----------------YTFA-------- 553

Qy        524 PITHILVCRPKRTKASMSEFLESEDGEV---EQQRTYSSGHNRLYFHS----DTCLPLRP 576
Db        554 ------MCSYREKKAEPQELLQLDGYTVDYTDPQPGLEGG--RAFFNAVKEGDTVI---- 601

Qy        577 QEMEVDSEDEKDP-EWLR---------EKTI--TQIEEFSDVNEG----EKEVMKLWNLH 620
Db        602 ----FASDDEQDRILWVQAMYRATGQSHKPVPPTQVQKLKPRAETCLSMDAPISQFYADR 657

Qy        621 VMKHG---FIADNQMN--HACMLFV-----------ENY--------GQKIIKKNLC-RN 655
Db        658 AQKHGMDEFISSNPCNFDHASLFEMVQRLTLDHRLNDSYSCLGWFSPGQVFVLDEYCARN 717

Qy        656 FMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASP 696
Db        718 -------------GVRGCHRHLCYLRDLLERAENGAMIDP 744
```

RESULT 4
A47392
chromodomain-helicase-DNA-binding protein, CHD-1 - mouse
N;Alternate names: KYBP protein
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Dec-2002
C;Accession: A47392; S21568
R;Delmas, V.; Stokes, D.G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 2414-2418, 1993
A;Title: A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI
A;Reference number: A47392; MUID:93211972; PMID:8460153
A;Accession: A47392
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1711 <DEL>
A;Experimental source: S194 plasmacytoma cells
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:128272, NCBIP:128273)
R;Delmas, V.; Perry, R.P.
submitted to the EMBL Data Library, May 1992
A;Description: KYBP, a mammalian protein that contains the SNF2/SWI2 helicase domain
A;Reference number: S21568
A;Accession: S21568
A;Molecule type: mRNA
A;Residues: 772-1711 <DE2>
A;Cross-references: EMBL:X66028
C;Superfamily: chromodomain helicase CHD1; chromobox homology
C;Keywords: DNA binding
F;293-336/Domain: chromobox homology <CB1>
F;387-427/Domain: chromobox homology <CB2>

  Query Match   4.0%; Score 153.5; DB 1; Length 1711;
  Best Local Similarity   20.1%; Pred. No. 0.27;
  Matches 124; Conservative   80; Mismatches 207; Indels 205; Gaps   31;

```
Qy          5 KHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAA 64
Db         11 RNGSGESSQSGDDCGSASGSGSGSSSGSSSDGSSSQS-GSSDSDSGSDSGSQSESESDTS 69
```

```
Qy         65 GAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSR 124
Db         70 RENKVQAKPPKVDGAE-----FWKSSPSILAVQRSAMLRKQPQQAQQQRPASSNSGSEED 124

Qy        125 TNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPS---PNSENEQNS 181
Db        125 SSSSEDS---DDSSSGAKRKKHNDEDWQMSG------------SGSPSQLGSDSESEE-- 167

Qy        182 VTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSN 241
Db        168 ------------ERDKSSCD----------------------------GTESDYEPKN 185

Qy        242 SHMVKSYSLLFRVTRPGRR--EFNG-MINGETNENIDV----------NEELPARRKRN- 287
Db        186 K---------VRSRKPQNRSKSKNGKKILGQKKRQIDSSEDEDDEDYDNDKRSSRRQATV 236

Qy        288 ----REDGE-KTFVAQMTVFDKNRRLQLL--------DGEYEVAMQEMEECPISKKRATW 334
Db        237 NVSYKEDEEMKT--------DSDDLLEVCGEDVPQPEDEEFE-TIERVMDCRVGRKGATG 287

Qy        335 ETI------LDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQE 388
Db        288 ATTTIYAVEADGDPNAGFERNKEPGDIQYLIKWKGWSHIHNT--------WETEETLKQQ 339

Qy        389 NKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN--ENRQKLRIFYQFLYNNNTRQQTEARD 446
Db        340 NVRGMKK------------LDNYKKKDQETKRWLKNASPEDVEY---YN----CQQELTD 380

Qy        447 DLHCPWCTLNCRKLYSLLKHLKLCHSR----------------FIFNYVYHPKGARIDVS 490
Db        381 DLH---------KQYQIVERI-IAHSNKQSAAGLPDYYCKWQGLPYSECSWEDGALISKK 430

Qy        491 INECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPI--THILVCRPKRTKASMSEFLESED 548
Db        431 FQTCID-EY--------------FSRN-QSKTTPFKDCKVLKQRPR--------FV---- 462

Qy        549 GEVEQQRTYSSGHNRL 564
Db        463 -ALKKQPSYIGGHEGL 477
```

RESULT 5
B32372
male-specific doublesex protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 02-Nov-1989 #sequence_revision 02-Nov-1989 #text_change 16-Feb-1997
C;Accession: B32372
R;Burtis, K.C.; Baker, B.S.
Cell 56, 997-1010, 1989
A;Title: Drosophila doublesex gene controls somatic sexual differentiation by producing
A;Reference number: A32372; MUID:89168451; PMID:2493994
A;Accession: B32372
A;Status: preliminary
A;Molecule type: DNA; mRNA
A;Residues: 1-549 <BUR>
A;Cross-references: GB:M25294; GB:J03156
C;Genetics:
A;Gene: FlyBase:dsx
A;Cross-references: FlyBase:FBgn0000504
C;Keywords: alternative splicing; DNA binding; transcription regulation

  Query Match 3.9%; Score 153; DB 2; Length 549;
  Best Local Similarity 27.5%; Pred. No. 0.063;
  Matches 47; Conservative 27; Mismatches 47; Indels 50; Gaps 5;

```
Qy         11 GGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSA---------------- 54
Db        242 GGSVGPATSSSGGGAPSSSNAAAATSSNGSSGGGGGGGGGSSGGGAGGGRSSGTSVITSA 301

Qy         55 ------------------SSSSSAAAAAGAAVLP----VKKPKMEHVQADHELFLQAFE 91
Db        302 DHHMTTVPTPAQSLEGSCDSSSPSPSSTSGAAILPISVSVNRKNGANVPLGQDVFLDYCQ 361

Qy         92 KPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVE 142
Db        362 KLLEKFRY--------PWEL-MPLMYVILKDADANIEEASRRIEE--ARVE 401
```

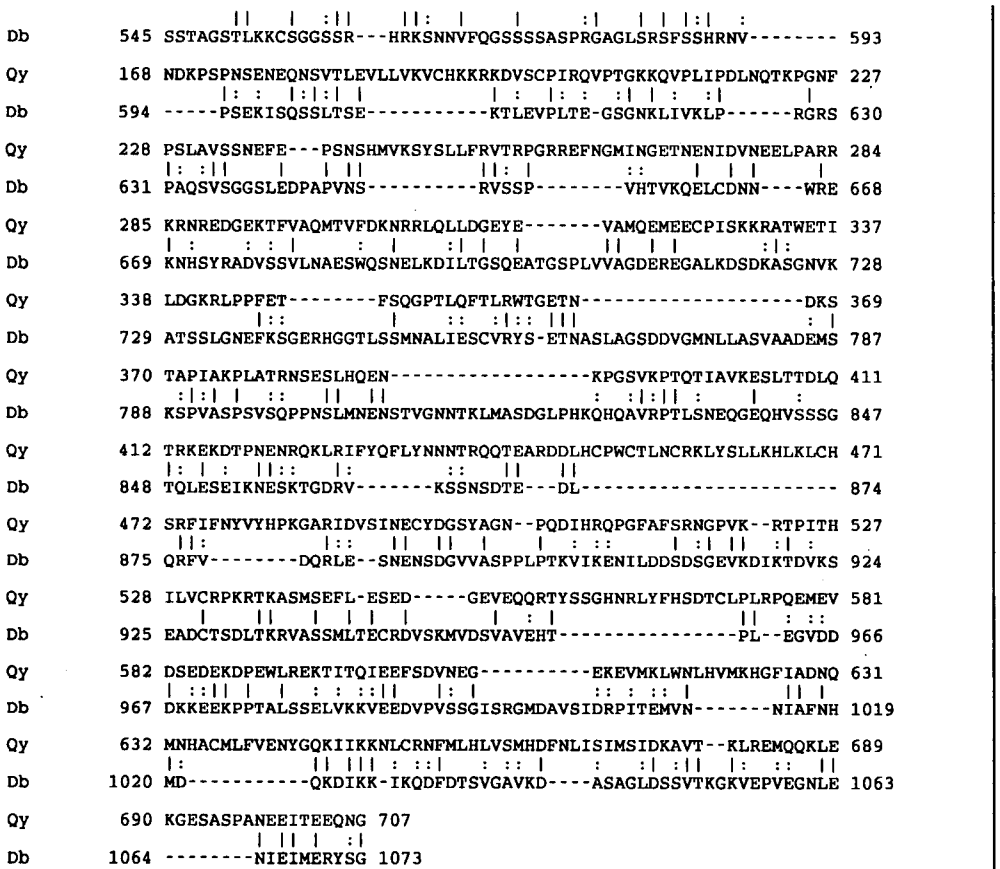
RESULT 6
KRMS2
keratin, type II cytoskeletal - mouse (fragment)
N;Alternate names: 67-kDa type II keratin
C;Species: Mus musculus (house mouse)
C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 20-Mar-1998
C;Accession: A02951
R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.
J. Biol. Chem. 260, 7142-7149, 1985
A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67
iate filament subunits.
A;Reference number: A92535; MUID:85207740; PMID:2581964
A;Accession: A02951
A;Molecule type: mRNA
A;Residues: 1-581 <STE>
A;Cross-references: GB:M10937; NID:g198622; PID:g387396
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil; intermediate filament
F;1-106/Domain: head (fragment) <HED>
F;1-70/Region: V1 subdomain (fragment)
F;71-106/Region: H1 subdomain
F;107-419/Domain: rod <ROD>
F;107-141/Region: coil 1A
F;142-153/Region: linker 1
F;154-254/Region: coil 1B
F;255-271/Region: linker 12
F;272-290/Region: coil 2A
F;291-298/Region: linker 2
F;299-419/Region: coil 2B
F;357/Region: stutter
F;420-581/Domain: tail <END>
F;420-439/Region: H2 subdomain
F;440-581/Region: V2 and E2 subdomains

  Query Match 3.9%; Score 153; DB 1; Length 581;
  Best Local Similarity 22.0%; Pred. No. 0.068;
  Matches 81; Conservative 53; Mismatches 140; Indels 94; Gaps 15;

```
Qy          6 HGGGG-GGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAA 64
Db         25 YGGGGFGGGSFGGGGFGGSGFGGGSGGGGGFGSGGGFGGGRFGGYGPVCSPSGIQEVTLN 84

Qy         65 GAAVLP-----------VKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPI---F 110
Db         85 QSLLQPLNVKVDPQIQKVKSQEREQIKSLNDKFASFIDK----VRFLEQQNKVLQTKWEL 140

Qy        111 LHRTLTYMSHRN------SRTNIKRKTFKVDDMLSKVEKMKGE-QESHSLSAHLQLTFTG 163
Db        141 LQQVDTTTRTQNLDPFFENYISILRR--KVDSLKSDQSRMESELKNMQDLVEEYRTKYED 198

Qy        164 FFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDV-SCPIRQVPTGKKQVPLIPDLNQT 222
Db        199 EINKR----TNAENE--FVTI----------KKDVDSAYMTKVELQAKRDALQQDIN-- 239

Qy        223 KPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVN----- 277
Db        240 ---------FFSTLYQMEMSQMQTQIS----------ETNVVLSMDNNRQFDLDGIISE 279

Qy        278 -----EELPARRKRNRE---------------DGEKTFVAQMTVFDKNRRLQLLDGEYE 316
Db        280 VKAQYDSICQRSKAEAETFYQSKYEELQITAGKHGDSVRNTKMEISELNRMIQRLRSEID 339

Qy        317 VAMQEMEE 324
Db        340 GCKKQISQ 347
```

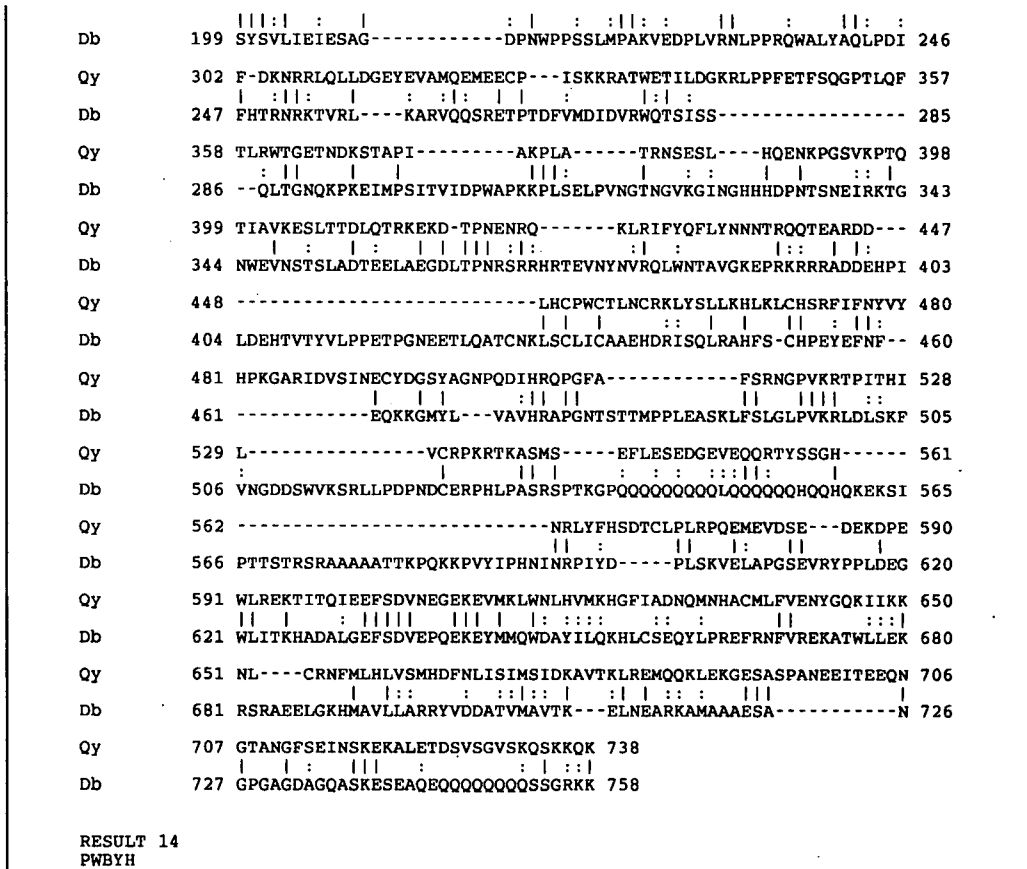
```
RESULT 7
A32372
female-specific doublesex protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 02-Nov-1989 #sequence_revision 02-Nov-1989 #text_change 23-Feb-1997
C;Accession: A32372
R;Burtis, K.C.; Baker, B.S.
Cell 56, 997-1010, 1989
A;Title: Drosophila doublesex gene controls somatic sexual differentiation by producing
A;Reference number: A32372; MUID:89168451; PMID:2493994
A;Accession: A32372
A;Status: preliminary
A;Molecule type: DNA; mRNA
A;Residues: 1-427 <BUR>
A;Cross-references: GB:M25292; GB:J03156
C;Genetics:
A;Gene: FlyBase:dsx
A;Cross-references: FlyBase:FBgn0000504
C;Keywords: alternative splicing; DNA binding; nucleus; transcription regulation

  Query Match             3.9%;  Score 152;  DB 2;  Length 427;
  Best Local Similarity   27.3%;  Pred. No. 0.052;
  Matches   45;  Conservative   25;  Mismatches   47;  Indels   48;  Gaps    4;

Qy         11 GGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSA---------------- 54
              ||  ||:  | |||   |:  ||||:| | ||||  |||||
Db        242 GGSVGPATSSSGGGAPSSSNAAAATSSNGSSGGGGGGGGGSSGGGAGGGRSSGTSVITSA 301

Qy         55 ------------------SSSSSAAAAAGAAVLP----VKKPKMEHVQADHELFLQAFE 91
                               ||| | :: :|||:||    | :    :|    ::||   :
Db        302 DHHMTTVPTPAQSLEGSCDSSSPSPSSTSGAAILPISVSVNRKNGANVPLGQDVFLDYCQ 361

Qy         92 KPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDD 136
              |  : :|:        |  |   | |:  :::  ||:  : ::::
Db        362 KLLEKFRY--------PWEL-MPLMYVILKDADANIEEASRRIEE 397


RESULT 8
T20268
hypothetical protein C56A3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20268
R;Sims, M.
submitted to the EMBL Data Library, July 1996
A;Reference number: Z19244
A;Accession: T20268
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-393 <WIL>
A;Cross-references: EMBL:Z77655; PIDN:CAB01137.1; GSPDB:GN00023; CESP:C56A3.1
A;Experimental source: clone C56A3
C;Genetics:
A;Gene: CESP:C56A3.1
A;Map position: 5
A;Introns: 51/3; 91/1; 121/1; 331/3

  Query Match             3.9%;  Score 151.5;  DB 2;  Length 393;
  Best Local Similarity   34.6%;  Pred. No. 0.05;
  Matches   46;  Conservative    8;  Mismatches   46;  Indels   33;  Gaps    5;

Qy          7 GGGGGGGSGPSAGSGGGGF--GGSAAVAAA--------------------TASGGKSGG 43
              |||||||  |    | ||||:  |||    |:|                      ||| | |
Db        117 GGGGGGGCGGGGGGGGGGYASGGSGGFASAPVSLPAPSYGGPPPPAPSFSHAPSGGYSSG 176

Qy         44 GSCGGGGSYSASSS----SSAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKP----TQ 95
              || ||| |   ||      :  ||||||    |   :      |       | |  |    ||
Db        177 GSSGGGYSSGGSSGGGGYAGGAAAAGATAAQV--DEASETSAGEPPVDQVFHSPKEPCTQ 234

Qy         96 IYRFLRTRNLIAP 108
                :::  |:   |
Db        235 QVKYIMLRSRKVP 247


RESULT 9
A56600
intermediate filament protein IF-1, cytosolic - common lancelet
C;Species: Branchiostoma lanceolatum (common lancelet)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 21-Jul-2000
C;Accession: A56600; S24265
R;Riemer, D.; Dodemont, H.; Weber, K.
Eur. J. Cell Biol. 58, 128-135, 1992
A;Title: Analysis of the cDNA and gene encoding a cytoplasmic intermediate filament (
amily.
A;Reference number: A56600; MUID:92354564; PMID:1644059
A;Accession: A56600
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-464 <RIE>
A;Cross-references: EMBL:X64522; NID:g5724; PIDN:CAA45827.1; PID:g5725
A;Note: sequence extracted from NCBI backbone (NCBIP:110465)
A;Note: intron positions were also determined
C;Superfamily: cytoskeletal keratin
C;Keywords: cytosol; intermediate filament

  Query Match             3.9%;  Score 151;  DB 2;  Length 464;
  Best Local Similarity   19.8%;  Pred. No. 0.067;
  Matches   95;  Conservative   77;  Mismatches  158;  Indels  150;  Gaps   20;

Qy          7 GGGGGGGSGPSAGSGGGGF-GGSAAVAAATASGGKSGGG--SCGGGGSYSASSSSSAAAA 63
              ||||||| | :: |  | | ||   : |   :  :| |   | ||||  :|  ::   |:
Db         18 GGGGGGGGGRASYSSSGRFSGGGGRMRAGGVTSRRSVGSSYSAGGGGRVNAVMAAGVMAS 77

Qy         64 AG---AAVLPVKKPKMEHVQADHEL------FLQAFEKPTQIYRFLRTRN--LIAPIFLH 112
              |    |:  :  :|       ||      |    :|     |:|: ||  | | | :
Db         78 VGGPAGALQTLSDARMTRAHEKQELSHLNDRFASYIDK----VRYLQERNSKLEAQIKIQ 133

Qy        113 RT--------------------LTYMSHRNSRTNIKRKTFKV----------------DD 136
               :                    :  :|:  ::  |:|  ::                :
Db        134 ESREAPNIKDLYEKELRDLRALVDELSNDKAQLEIERNNWQEQAEDYKMKWQDEAGLRSE 193

Qy        137 MLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
              : :::|::| | :: :: | | |                 ||: ::   |:  :|  | :
Db        194 LEAEIERLKKELDAATM-ARLDL---------------ENKLSTAQEEIDFLKRVHDEE 236

Qy        197 KDVSCPIRQVPTGKKQVPLIPDLNQTKPGNF---PSLAVSSNEFEPSNSHMVKSYSLLFR 253
                   |||:      :   | :::       |   | |  :  |             | | |
Db        237 ------IRQLQDQLNESLTIVEVDSRAASTFAPGPDLTEALREIR-------TQYEELGR 283

Qy        254 VTRPGRREFNGMINGETNENIDVN-----EELPARRKRNREDGEKTFVAQMTVFDKNR-- 306
              | |               |: ||       ||  :|:|   | |    |:  | :  |
Db        284 VNR---------------EDADVKYKQKFSELAHQRER---DNEALMTARTEVTELRRNL 325

Qy        307 ------------------------RLQLLDGEYEVAMQEMEE------CPISKKRATW 334
                                      |:||   ||: |::::|:       ::::   :
Db        326 NSLVAENEALKSKNHALEGSLAELEARMQLEIEEYQAAIRDLEQELETKTSEMAQQMQAY 385

Qy        335 ETILDGKRLPPFETFSQGPTLQ-FTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGS 393
              : ::| |     |  :    |:  :|  ||: :       :  :: :| |     | ||
Db        386 KMLMDTKMALDMEIAAYRKLLEGEEIRLFGESKEG-----VQQTSSSSSSSYQYSMKSGS 440


RESULT 10
T31328
fibroin - Chinese oak silkmoth
C;Species: Antheraea pernyi (Chinese oak silkmoth)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31328
R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A;Description: Characterization of the full length fibroin gene of a wild silkworm, A
A;Reference number: Z20995
A;Accession: T31328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2639 <SEZ>
```

```
A;Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C;Genetics:
A;Introns: 14/3

  Query Match              3.8%;  Score 149.5;  DB 2;  Length 2639;
  Best Local Similarity   43.6%;  Pred. No. 0.85;
  Matches   34;  Conservative    9;  Mismatches   20;  Indels   15;  Gaps    2;

Qy          5 KHGGG----------GGGGSGPSAGSG----GGGFGGSAAVAAATASGGKSGGGSCGGG 49
              : |||           | | |  :||||    |||:|  :| ||| |:   || |  ||
Db        134 RRGGGFYETHNSYSSYGSGSSSAAAGSGAGGVGGGYGSDSAAAAAAAAAAASGAGGSGGY 193

Qy         50 GSYSASSSSSAAAAAGAA 67
              | | : |:::|||| ||
Db        194 GGYGSDSAAAAAAAAAAA 211


RESULT 11
KRHU2
keratin 1, type II, cytoskeletal - human
N;Alternate names: 67K type II epidermal keratin; cytokeratin 1
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1986 #sequence_revision 22-Oct-1999 #text_change 10-Dec-1999
C;Accession: A22940; A02950; A43342
R;Johnson, L.D.; Idler, W.W.; Zhou, X.M.; Roop, D.R.; Steinert, P.M.
Proc. Natl. Acad. Sci. U.S.A. 82, 1896-1900, 1985
A;Reference number: A22940; MUID:85166239; PMID:2580302
A;Accession: A22940
A;Molecule type: DNA
A;Residues: 1-643 <JOH>
A;Cross-references: GB:M98776; GB:M11215; GB:M11845; GB:M11846; NID:g1843461
A;Note: translation of initiator Met is not shown
R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.
J. Biol. Chem. 260, 7142-7149, 1985
A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,00
iate filament subunits.
A;Reference number: A92535; MUID:85207740; PMID:2581964
A;Accession: A02950
A;Molecule type: mRNA
A;Residues: 151-183,'K',185-199,'M',201-204,'K',206-236,'S',238-239,'R',241-356,'Y',358-
,'S',638-643 <STE>
A;Cross-references: GB:M10938; NID:g186787; PIDN:AAA36153.1; PID:g386854
A;Experimental source: tissue neonatal foreskin
A;Note: the authors translated the codon CUG for residue 476 as Met
R;Chipev, C.C.; Korge, B.P.; Markova, N.; Bale, S.J.; DiGiovanna, J.J.; Compton, J.G.; S
Cell 70, 821-828, 1992
A;Title: A leucine----proline mutation in the H1 subdomain of keratin 1 causes epidermol
A;Reference number: A43342; MUID:92386601; PMID:1381288
A;Accession: A43342
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 144-146,'P',148-159,'P',161-183,'K',185-186 <CHI>
A;Cross-references: GB:M98776; GB:M11215; GB:M11845; GB:M11846; NID:g1843461
A;Note: sequence extracted from NCBI backbone (NCBIP:112784)
C;Comment: The cytoskeletal and microfibrillar keratins are classified into two types, t
atin IF protein subunit appears to be a heterotetramer of two type I and two type II pro
C;Comment: Keratin 1 is expressed in terminally differentiating epidermis.
C;Genetics:
A;Gene: GDB:KRT1
A;Cross-references: GDB:128198; OMIM:139350
A;Map position: 12q11-12q13
A;Note: defects in this gene may result in epidermolytic hyperkeratosis
C;Complex: heterotetramer of two type I, usually keratin 10 (see PIR:KRHU0), and two typ
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil; heterotetramer; intermediate filament
F;4-179/Domain: head <HED>
F;4-143/Region: E1 and V1 subdomains
F;14-179/Region: H1 subdomain
F;180-492/Domain: rod <ROD>
F;180-214/Region: coil 1A
F;215-226/Region: linker 1
F;227-327/Region: coil 1B
F;328-344/Region: linker 12
F;345-363/Region: coil 2A
F;364-371/Region: linker 2
F;372-492/Region: coil 2B
F;430/Region: stutter
F;493-643/Domain: tail <END>
F;493-512/Region: H2 subdomain
F;513-643/Region: V2 and E2 subdomains

  Query Match              3.8%;  Score 148.5;  DB 1;  Length 643;
  Best Local Similarity   20.5%;  Pred. No. 0.15;
  Matches   83;  Conservative   71;  Mismatches  167;  Indels   83;  Gaps   16;

Qy          7 GGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGG--------GGSYSASSSS 58
              ||| |||     | |||||||  :       ||  |||  ||         ||    : :
Db         99 GGGFGGGGFGGGGIGGGGFGGFGSGGGGFGGGGFGGGGGYGGGYGPVCSPGGIQEVTINQ 158

Qy         59 SAAAAAGAAVLP----VKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRT 114
              |       : |    ||  : | :|: :  |    :|     |||  :| :  :
Db        159 SLLQPLNVEIDPEIQKVKSREREQIQSLNNQFASFIDK----VRFLEQQNQV--LQTKWE 212

Qy        115 LTYMSHRNSRTNIKRKTFK--VDDMLSKVEKMKGEQES-HSLSAHLQLTFTGFFHKND-- 169
              |       ::||:      |:  ::::   |:::| :|    |   ::|     : :| :
Db        213 LLQQVDTSTRTHNLEPYFESFINNLRRGVDQLKSDQSRLDSELKNMQDMVEDYRNKYEDE 272

Qy        170 -KPSPNSENEQNSVTLEV---LLVKVCHKKRKD-----------VSCPIRQVPTGKKQV 213
               |:|||  ::  :|    : ||  : : |                    : |: |   :
Db        273 INKRTNAENEFVTIKKDVDGAYMTKVDLQAKLDNLQQEIDFLTALYQAELSQMQTQISET 332

Qy        214 PLIPDLNQTKPGNFPSL----------AVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFN 263
              :|  ::  :  :  |:              :: |  : :  |   |     | |
Db        333 NVILSMDNNRQFDLDSIIAEVKAQNEDIAQKSKAEAESLYQSKYEELQITAGRHG----- 387

Qy        264 GMINGETNENIDVNE-ELPARRKRNREDGEKTFVA--QMTVFDKNRRLQ--LLD-----G 313
              :   |  |:::|     :| |:  |  |  ::  | :: |  :| :  | |
Db        388 ---DSVRNSKIEISELNRVIQRLRSEIDNVKKQISNLQQSISDAEQRGENALKDAKNKLN 444

Qy        314 EYEVAMQEMEE---------------CPISKKRATWETILDGK 341
              : | |:|: :|               :  : ||: |:|:|:
Db        445 DLEDALQQAKEDLTRLLRDYQELMNTKLALDLEIATYRTLLEGE 488


RESULT 12
T06677
hypothetical protein T17F15.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06677
R;Quetier, F.; Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Art
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15793
A;Accession: T06677
A;Molecule type: DNA
A;Residues: 1-1611 <QUE>
A;Cross-references: EMBL:AL049658; GSPDB:GN00061; ATSP:T17F15.70
A;Experimental source: cultivar Columbia; BAC clone T17F15
C;Genetics:
A;Gene: ATSP:T17F15.70
A;Map position: 3
A;Introns: 46/3; 170/3

  Query Match              3.8%;  Score 148.5;  DB 2;  Length 1611;
  Best Local Similarity   20.1%;  Pred. No. 0.51;
  Matches  160;  Conservative  100;  Mismatches  287;  Indels  251;  Gaps   36;

Qy          6 HGGGGGGGSGPSAGSGGGGFGGSAAVA----------AATASGGKSGGGSCGGGGSYSA 54
              |||   |||  :  :       | :|:           ||: |      | | ||: |
Db        431 HGGRHSGGSAEANKTSSSHLHASKSVSVKQQVENNLKCVATSPGSTRSAPSPGSGGNVSK 490

Qy         55 SSSSSAAAAAGAA--VLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIA----- 107
                | | | :  :  ||  |       |          :      :| ||
Db        491 DGQQRNAGAGGVSEVLAAVKDEKSSSSSQSHN------NSQSCSSEHAKTGNLCGKEDAR 544

Qy        108 PIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHK 167
```



RESULT 13
T51002
hypothetical protein B7J19.10 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51002
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51002
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-863 <SCH>
A;Cross-references: EMBL:AL389899; GSPDB:GN00116; NCSP:B7J19.10
A;Experimental source: BAC clone B7J19; strain OR74A
C;Genetics:
A;Gene: NCSP:B7J19.10
A;Map position: 6
A;Introns: 782/3

Query Match 3.8%; Score 148; DB 2; Length 863;
Best Local Similarity 18.0%; Pred. No. 0.24;
Matches 114; Conservative 78; Mismatches 228; Indels 212; Gaps 25;



RESULT 14
PWBYH
potassium transport protein TRK1, high-affinity - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J0693; protein YJL129c
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 12-Nov-1999
C;Accession: S05849; S56910
R;Gaber, R.F.; Styles, C.A.; Fink, G.R.
Mol. Cell. Biol. 8, 2848-2859, 1988
A;Title: TRK1 encodes a plasma membrane protein required for high-affinity potassium
A;Reference number: S05849; MUID:88302204; PMID:3043197
A;Accession: S05849
A;Molecule type: DNA
A;Residues: 1-1235 <GAB>
A;Cross-references: EMBL:M21328; NID:g171803; PIDN:AAA34728.1; PID:g171804
R;Czieplu ch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56891
A;Accession: S56910
A;Molecule type: DNA
A;Residues: 1-1235 <CZI>
A;Cross-references: EMBL:Z49404; NID:g1008329; PIDN:CAA89424.1; PID:g1008330; GSPDB:G
C;Genetics:
A;Gene: SGD:TRK1; MIPS:YJL129c
A;Cross-references: SGD:S0003665; MIPS:YJL129c
A;Map position: 10L
C;Superfamily: potassium transport protein TRK1
C;Keywords: ion transport; potassium transport; transmembrane protein
F;50-70/Domain: transmembrane #status predicted <TM01>

```
F;72-96/Domain: transmembrane #status predicted <TM02>
F;107-127/Domain: transmembrane #status predicted <TM03>
F;778-800/Domain: transmembrane #status predicted <TM04>
F;812-834/Domain: transmembrane #status predicted <TM05>
F;839-859/Domain: transmembrane #status predicted <TM06>
F;862-882/Domain: transmembrane #status predicted <TM07>
F;898-918/Domain: transmembrane #status predicted <TM08>
F;923-943/Domain: transmembrane #status predicted <TM09>
F;972-991/Domain: transmembrane #status predicted <TM10>
F;1079-1096/Domain: transmembrane #status predicted <TM11>
F;1111-1132/Domain: transmembrane #status predicted <TM12>
F;738/Binding site: ATP (Lys) #status predicted

  Query Match             3.8%;  Score 146.5;  DB 1;  Length 1235;
  Best Local Similarity   20.1%;  Pred. No. 0.47;
  Matches  154;  Conservative   90;  Mismatches  227;  Indels  295;  Gaps   39;

Qy         97 YRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKV----------EKMK 145
              ::  ||: ::      ||:|       ||:  ||  | ||   |:           : :
Db        148 FKMRRTKTILERELTARTMTKNRTGTQRTSYPRKQAKTDDFQEKLFSGEMVNRDEQDSVH 207

Qy        146 GEQESHSLSAHLQLTFTGF---------FHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
               :| || :|      |            | | |:   | | ::|:
Db        208 SDQNSHDISRDSSNNNTNHNGSSGSLDDFVKEDETDDNGEYQENN-------------- 252

Qy        197 KDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEF-EPSNSHMVKSYSLLF--- 252
                                      :: ::  |||   : | :   |  || |
Db        253 ----------------------------SYSTVGSSSNTVADESLNQKPKPSSLRFDEP 283

Qy        253 -RVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLL 311
               || |               |  |  |:|: :|:          ::  |  | : :|
Db        284 HSKQRPAR---------------VPSEKFAKRRGSRD---------ISPADMYRSIMML 318

Qy        312 DGEYEVAMQEMEECPI---------------SKKRATWETILDGKRLPPFETFSQGPTL 355
               |::| |  | |  |:                   |:||    ::| ::   :  :: .|
Db        319 QGKHE-ATAEDEGPPLVIGSPADGTRYKSNVNKLKKATG---INGNKIKIRDKGNESNTD 374

Qy        356 QFTLRWTGETNDKSTAPIAKPLATRNSESLHQE--NKPGSVKPTQT-------IAVKESL 406
               | ::  : | |  ||| ::      :  |||   ||  |:: | |       ||: ::
Db        375 QNSV--SSEAN--STASVS------DESSLHTNFGNKVPSLR-TNTHRSNSGPIAITDNA 423

Qy        407 TTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTR---QQTEARDDLHCPWCTLNCRKLYS- 462
               ||     :|  |:          ||      |   ::    |||: |  ::  ||  |
Db        424 ETD-----KKHGPS---------IQFDITKPPRKISKRVSTFDDLN-PKSSVLYRKKASK 468

Qy        463 --LLKHLKLCHSRFIFNYVYHPKGARIDVSI-NECYDGSYAGNPQDIHRQPGFAFSRNGP 519
                 |:||                ||  ||   |      ||   |            | |
Db        469 KYLMKHF-------------PKARRIRQQIKRRLSTGSIEKN-----------SSNNV 502

Qy        520 VKRTPITHILVCRPKRTKASMSE--------------FLESEDGEVEQQRTYSSGHNRL 564
               | |||              | :              | ::| |: | :|   |
Db        503 SDRKPIT-----------DMDDDDDDDDNDGDNNEEYFADNESGD-EDERVQQSEP--- 546

Qy        565 YFHSDTCLPLRPQEME---------------VDSEDEKDPEWLREKTITQIEEFSDVNE 608
               |||: |    |: |                |    :     | :||  : |  |:||
Db        547 --HSDSELKSHQQQQEKHQLQQNLHRMYKTKSFDDNRSRAVPMERSRTI-DMAEAKDLNE 603

Qy        609 ------GEKEVMKLWNLH--------------VMKHG-FIADNQMNHACMLFVENYGQK 646
                    :| | : |  |                : :|| : :|:  |:        | |
Db        604 LARTPDFQKMVYQNWKAHHRKKPNFRKRGWNNKIFEHGPYASDSDRNYPDN---SNTG-- 658

Qy        647 IIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQN 706
                  | :||          ||:  | :          |    |::| :|| |
Db        659 --------NSILHYAE-------SILHHDGS--------HKNGSEEASSDSNENIYSTNG 695

Qy        707 GT----ANGFSEINSKEKA---------LETDSVSGVSKQSKKQKL 739
               |:     | :   |  |:         : |    :|| | |  |
Db        696 GSDHNGLNNYPTYNDDEEGYYGLHFDTDYDLDPRHDLSKGSGKTYL 741


RESULT 15
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st

N;Alternate names: gp195 surface antigen
C;Species: Plasmodium falciparum
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C;Accession: S05603; S04850
R;Myler, P.J.
submitted to the EMBL Data Library, April 1989
A;Reference number: S05603
A;Accession: S05603
A;Molecule type: mRNA
A;Residues: 1-1639 <MYL>
A;Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R;Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Pl
A;Reference number: S04850; MUID:89345116; PMID:2668887
A;Accession: S04850
A;Molecule type: mRNA
A;Residues: 1504-1639 <MYL2>
A;Cross-references: EMBL:X15063
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

  Query Match             3.8%;  Score 146.5;  DB 2;  Length 1639;
  Best Local Similarity   20.6%;  Pred. No. 0.69;
  Matches  168;  Conservative  123;  Mismatches  291;  Indels  235;  Gaps   46;

Qy         10 GGGGSGPSAGSGGG-GFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAV 68
               |  ||  | ||||    ||| |   : ||||    |  |     : | :|| : |   |
Db         68 GSKGSVASGGSGGSVASGGSVASGGSVASGGSVASGGSGNSRRTNPSDNSSDSDAKSYA- 126

Qy         69 LPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIK 128
                        ::|   :: | ::  : | |::    | :::       ||         ||
Db        127 ------DLKHRVRNYLLTIKELKYP-QLFDL--TNHML-------TLC--------DNIH 162

Qy        129 RKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLL 188
                 : :|          | :| : |   |  |                       |  |
Db        163 GFKYLID----------GYEEINELLYKLNFYFD-------------------LLRAKL 192

Qy        189 VKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSY 248
                ||  '   :  |           |:|               :| : :|| :     ::|
Db        193 NDVC----ANDYC----------QIPF-------------NLKIRANELD-----VLK-- 218

Qy        249 SLLFRVTRP--------GRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
               |:|  :|         |: |      | :| |||  || :         |: :||
Db        219 KLVFGYRKPLDNIKDNVGKMEDYIKKNKKTIENI--NELI--------EESKKTIDKNKN 268

Qy        301 VFDKNRRLQLLDGEYEVAM--QEMEEC-----------PISKKRATWETILDG----KRL 343
                 :  : :|   :|::::  :::||                ||    : :||     |
Db        269 ATKEEEKKKLYQAQYDLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNP 328

Qy        344 PPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSES----------LHQENKPG 392
               ||  :    |         | ::|    ||| :   | :|           | ::||
Db        329 PPANS-GNTPNTLLDKNKKIEEHEKEIKEIAKTIKF-NIDSLFTDPLELEYYLREKNK-- 384

Qy        393 SVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN--NNTRQQTEARDDLHC 450
                    | :  : |     ||   |||      : |   ||  ||   :  :  ||
Db        385 ------NIDISAKVET-----KESTEPNEYPN--GVTYPLSYNDINNALNELNSFGDLIN 431

Qy        451 PW----------CTLNCRKLY--SLLKHLKLCHSRFIFN-YVYHPKGARI-DVS------ 490
               |:           | | || :   : : :|:   :   :   |  :   : |::
Db        432 PFDYTKEPSKNIYTDNERKKFINEIKEKIKIEKKKIESDKKSYEDRSKSLNDITKEYEKL 491

Qy        491 INECYDGSYAGNPQ--DIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEF---LE 545
               :|| ||  :  |    :  :  |  :|      |   :||         | ||        ||
Db        492 LNEIYDSKFNNNIDLTNFEKMMGKRYS----YKVEKLTH------HNTFASYENSKHNLE 541

Qy        546 SEDGEVEQQRTYSSGH----NRLYFHSDTCLPLRPQEMEVDSED-EKDPEWLREKTITQI 600
                   ::    ||  :       | :: :    :   |:|    |: :|| | | || ||:
Db        542 KLTKALKYMEDYSLRNIVVEKELKYYKNLISKIE-NEIETLVENIKKDEEQLFEKKITK- 599

Qy        601 EEFSDVNEGEKEVMKLWNL-HVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLH 659
```

```
                 | |: ::::::: ::  |     :  |:::       |    ::| ||:     :|
Db        600 ----DENKPDEKILEVSDIVKVQVQKVLLMNKID-------ELKKTQLILKNVELKHNIH 648

Qy        660 LVSMH-------DFNLISIMSIDKAVTKLREMQQKLE----------KGESASPANEEIT 702
              : : :         : ||    : | : ||:    |:|          | |  |  :|  |
Db        649 VPNSYKQENKQEPYYLI---VLKKEIDKLKVFMPKVESLINEEKKNIKTEGQSDNSEPST 705

Qy        703 EEQNGTANGFSEINSKEK---ALETDSVSGVSKQSKK 736
              |   |   | :     ::   ||| |||   ::: |:
Db        706 E---GEITGQATTKPGQQAGSALEGDSVQAQAQEQKQ 739

Search completed: September  2, 2003, 11:02:54
Job time : 29.834 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37 ; Search time 23.3289 Seconds
(without alignments)
1340.299 Million cell updates/sec

Title: US-09-874-162A-5
Perfect score: 3885
Sequence: 1 MAPQKHGGGGGGGSGPSAGS.....KALETDSVSGVSKQSKKQKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	160	4.1	3969	3	US-08-061-376-5	Sequence 5, Appli
2	153	3.9	738	3	US-08-864-038A-3	Sequence 3, Appli
3	147	3.8	832	1	US-08-209-747-2	Sequence 2, Appli
4	147	3.8	832	1	US-08-458-298-2	Sequence 2, Appli
5	146.5	3.8	1235	1	US-08-118-101A-2	Sequence 2, Appli
6	142.5	3.7	474	4	US-09-702-705-1812	Sequence 1812, Ap
7	142.5	3.7	474	4	US-09-736-457-1812	Sequence 1812, Ap
8	142.5	3.7	712	4	US-09-059-584-51	Sequence 51, Appl
9	141	3.6	201	3	US-09-052-995-1	Sequence 1, Appli
10	141	3.6	201	3	US-09-053-003-40	Sequence 40, Appl
11	141	3.6	201	4	US-09-054-281-22	Sequence 22, Appl
12	141	3.6	201	4	US-09-478-948-6	Sequence 6, Appli
13	139.5	3.6	912	1	US-07-789-915A-8	Sequence 8, Appli
14	139.5	3.6	912	1	US-08-005-002C-8	Sequence 8, Appli
15	139.5	3.6	912	1	US-08-487-203A-8	Sequence 8, Appli
16	139	3.6	100	4	US-09-411-067C-4	Sequence 4, Appli
17	137.5	3.5	647	2	US-08-770-761A-8	Sequence 8, Appli
18	137.5	3.5	705	2	US-08-770-761A-7	Sequence 7, Appli
19	137	3.5	235	2	US-08-529-190B-1	Sequence 1, Appli
20	137	3.5	641	4	US-09-249-585A-3	Sequence 3, Appli
21	137	3.5	641	4	US-09-410-399-4	Sequence 4, Appli
22	134.5	3.5	689	4	US-08-778-570B-16	Sequence 16, Appl
23	134.5	3.5	689	4	US-09-059-584-16	Sequence 16, Appl
24	134.5	3.5	709	4	US-08-778-570B-15	Sequence 15, Appl
25	134.5	3.5	709	4	US-09-059-584-15	Sequence 15, Appl
26	131	3.4	2211	4	US-09-738-884-1	Sequence 1, Appli
27	129.5	3.3	1332	1	US-07-609-716-41	Sequence 41, Appl
28	129.5	3.3	1332	3	US-08-475-411A-41	Sequence 41, Appl
29	129.5	3.3	1332	3	US-08-478-029A-41	Sequence 41, Appl
30	129.5	3.3	1494	3	US-08-755-587-186	Sequence 186, App
31	129	3.3	447	4	US-09-916-109-5	Sequence 5, Appli
32	129	3.3	483	4	US-09-916-109-4	Sequence 4, Appli
33	127.5	3.3	146	1	US-07-609-716-105	Sequence 105, App
34	127.5	3.3	146	3	US-08-475-411A-105	Sequence 105, App
35	127.5	3.3	146	3	US-08-478-029A-105	Sequence 105, App
36	127.5	3.3	609	4	US-09-107-532A-4451	Sequence 4451, Ap
37	127.5	3.3	1345	2	US-08-977-767-3	Sequence 3, Appli
38	127	3.3	651	3	US-08-556-978B-19	Sequence 19, Appl
39	127	3.3	651	3	US-09-247-806-1	Sequence 1, Appli
40	127	3.3	718	1	US-08-425-069-2	Sequence 2, Appli
41	127	3.3	718	2	US-08-317-844B-2	Sequence 2, Appli
42	127	3.3	747	3	US-09-034-177-3	Sequence 3, Appli
43	126.5	3.3	40	3	US-08-815-190A-17	Sequence 17, Appl
44	126	3.2	123	4	US-09-072-596-243	Sequence 243, App
45	126	3.2	142	1	US-07-609-716-100	Sequence 100, App

ALIGNMENTS

RESULT 1
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabali, Malek
; APPLICANT: Selleri, Licia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-061-376-5

Query Match 4.1%; Score 160; DB 3; Length 3969;
Best Local Similarity 20.6%; Pred. No. 0.0008;
Matches 181; Conservative 97; Mismatches 308; Indels 292; Gaps 40;

Qy 3 PQKHGGGGGGG----------SGPSAGSGGGGFGGS-AAVAAATASGGKSGG 43

```
Db         13 PGTTGGGGGGGRRGLGGDPRQRVPALLLPPGPPVGGGGPGAPPSPPAVAAAAAAAGSSGA 72

Qy         44 GSCGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTR 103
              |  ||  : ||:|||||:::: ::      | :  |    :  ||
Db         73 GVPGGAAAASAASSSSASSSSSSSSSASSGPALLRVGPGFDAALQ-------------- 117

Qy        104 NLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTG 163
               : | |               ||::|  |:     |      || |          | |
Db        118 -VSAAI-------------GTNLRR--FRAVFGESGGGGGSGEDEQ----------FLG 150

Qy        164 FFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTK 223
              |         |: |          |:|              : ||    |   |   : :
Db        151 F---------GSDEE----------VRV------------RSPTRSPSVKTSPRKPRGR 178

Qy        224 P--GNFPSLAVSSNE--FEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEE 279
              |  |:  : |: |:   | | |    ||                        : |   :
Db        179 PRSGSDRNSAILSDPSVFSPLNKSETKS----------------------GDKIKKKDS 215

Qy        280 LPARRKRNREDGEKTFVAQMTVFDKNRRL-QLLDGEYEVAMQEMEECP---------ISK 329
                :|| |     ||          : : :|  |  | ::::::  |          | |
Db        216 KSIEKKRGR---PPTFPGVKIKITHGKDISELPKGNKEDSLKKIKRTPSATFQQATKIKK 272

Qy        330 KRATWETILDGKRLPPFETFSQGPTLQFTLRWTGET------NDKSTAPIAKPLATRNSE 383
               ||       ||  |   :  ||   :    :  :|      | |   |    |   | :
Db        273 LRA-------GKLSPSSLSLRQGSFKZEGRGYKLZTERKASINRKDKDPFGL-LIILNWK 324

Qy        384 SLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIF------YQFLYNNN 437
              |     | |  :        || |    : |:  |   |   :| :|        | | :|:
Db        325 S---PRKSGKTR-------KEHLHLQKKIRQLSDKALEGLSQLGLFLLQKGQMQPLLSNS 374

Qy        438 TR-QQTEARDDLHCPWCTLNCRKLYSLLKHLK---------------LCHSRFIFNYVYH 481
               | |:  :|  :      |  ||: : :|:::                    |||  :  |
Db        375 YRGQKKGAQKKIEKEAAQLQGRKVKTQVKNIRQFIMPVVSAISSRIIKTPRRFIEDEDYD 434

Qy        482 P--KGARIDVSINECYDGSYAGNPQ--------------DIHR--QPGFAFSRNG----- 518
              |  | ||:: : |  :     |: :                |  |   |    | :
Db        435 PPIKIARLESTPNSRFSAPSCGSSEKSSAASQHSSQMSSDSSRSSSPSVDTSTDSQASEE 494

Qy        519 ------------------PVKRTPITHILVCRPKRTKASMSEF----------LESEDGE 550
                              |: ::|       | :|   |   |          |:|
Db        495 IQVLPEERSDTPEVHPPLPISQSPENESNDRRSRRYSVSERSFGSRTTKKLSTLQS---- 550

Qy        551 VEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEF-----SD 605
               ||:| ||    |        ||:|      |  :  | ||   ||     |      :
Db        551 APQQQTSSSPPPPLLTPPP---PLQP----ASSISDHTP-WLMPPTIPLASPFLPASTAP 602

Qy        606 VNEGEKEVMK----LWNLHVMKHGFIADNQMNHACMLFVENYGQK-IIKKNLCRNFMLHL 660
              :   | :::      |   :||      : |       | :: :|:| :  ||
Db        603 MQGKRKSILREPTFRWT--SLKHSRSEPQYFSSA------KYAKEGLIRKPIFDNFRPPP 654

Qy        661 VSMHDFNLISIMSIDKAVTKLR-----------EMQQ--------KLEKGESASPANEEI 701
              ::  |    |  |        |           :| :        :    |: |   | :
Db        655 LTPEDVGFASGFSASGTAASARLFSPLHSGTRFDMHKRSPLLRAPRFTPSEAHSRIFESV 714

Qy        702 TEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              |   | |: | |                |||| : :|:|:
Db        715 TLPSNRTSAGTSS------------SGVSNRKRKRKV 739


RESULT 2
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
;    FEATURE:
;      NAME/KEY:  peptide
;      LOCATION:  from 1 to 738
;      IDENTIFICATION METHOD:  E (by experiment)
US-08-864-038A-3

  Query Match             3.9%;  Score 153;  DB 3;  Length 738;
  Best Local Similarity   43.5%;  Pred. No. 0.00024;
  Matches   37;  Conservative   11;  Mismatches   11;  Indels   26;  Gaps    3;

Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGG-GSCGGGGSYSAS---------- 55
              |||||||:|   |:|||| ||: |:||| |: |  || |  ||||: :|:
Db        437 GGGGGGGNG--GGNGGGGGGGAGALAAALAAAGAGGGLGGGGGGGALAAALAAAGAGGGG 494

Qy         56 ------------SSSSAAAAAGAA 67
                          |:::||||| ||
Db        495 FGGLGGLGGLGGGSAAAAAAAAAAA 519


RESULT 3
US-08-209-747-2
; Sequence 2, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  N. clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..309
US-08-209-747-2

  Query Match             3.8%;  Score 147;  DB 1;  Length 832;
  Best Local Similarity   37.4%;  Pred. No. 0.00094;
  Matches   34;  Conservative    9;  Mismatches   18;  Indels   30;  Gaps    2;

Qy          6 HGGGGGGGSGPSAGSGGGGFGG----------SAAVAAATASGGK--------------- 40
              :| | | |:| :||:| ||:||           || |||| :||
Db        424 YGRGAGAGAGAAAGAGAGGYGGQGGYGAGAGAGAAAAAATGAGGAGGYGRGAGAGAGAAA 483

Qy         41 -----SGGGSCGGGGSYSASSSSSAAAAAGA 66
                   :||   || | | | : : |||||||
Db        484 GAGAGTGGAGYGGQGGYGAGAGAGAAAAAGA 514


RESULT 4
US-08-458-298-2
; Sequence 2, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  N. clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..309
US-08-458-298-2

  Query Match             3.8%;  Score 147;  DB 1;  Length 832;
  Best Local Similarity   37.4%;  Pred. No. 0.00094;
  Matches   34;  Conservative    9;  Mismatches   18;  Indels   30;  Gaps    2;

Qy          6 HGGGGGGGSGPSAGSGGGGFGG----------SAAVAAATASGGK--------------- 40
              :| | | |:| :||:| ||:||           || |||| :||
Db        424 YGRGAGAGAGAAAGAGAGGYGGQGGYGAGAGAGAAAAAATGAGGAGGYGRGAGAGAGAAA 483

Qy         41 -----SGGGSCGGGGSYSASSSSSAAAAAGA 66
                   :||   || | | | : : |||||||
Db        484 GAGAGTGGAGYGGQGGYGAGAGAGAAAAAGA 514


RESULT 5
US-08-118-101A-2
; Sequence 2, Application US/08118101A
; Patent No. 5620892
;  GENERAL INFORMATION:
;    APPLICANT:  Kurtz, Stephen E.
;    APPLICANT:  Knickerbocker, Aron M.
;    APPLICANT:  McCullough, John R.
;    TITLE OF INVENTION:  A STRAIN OF SACCHAROMYCES CEREVISIAE
;    TITLE OF INVENTION:  EXPRESSING THE GENE ENCODING POTASSIUM TRANSPORTER MINK
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burton Rodney
;      STREET:  P.O. Box 4000
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  08543-4000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/118,101A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaul, Timothy J.
;      REGISTRATION NUMBER:  33,111
;      REFERENCE/DOCKET NUMBER:  DC27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 252-5901
;      TELEFAX:  (609) 252-4526
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1235 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-118-101A-2

  Query Match             3.8%;  Score 146.5;  DB 1;  Length 1235;
  Best Local Similarity   20.1%;  Pred. No. 0.0019;
  Matches  154;  Conservative   90;  Mismatches  227;  Indels  295;  Gaps   39;

Qy          97 YRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKV----------EKMK 145
               ::  ||: ::      ||:|       ||:  ||  | ||   |:           : :
Db         148 FKMRRTKTILERELTARTMTKNRTGTQRTSYPRKQAKTDDFQEKLFSGEMVNRDEQDSVH 207

Qy         146 GEQESHSLSAHLQLTFTGF---------FHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
               :| || :|       |           | | |:   | | ::|:
Db         208 SDQNSHDISRDSSNNNTNHNGSSGSLDDFVKEDETDDNGEYQENN-------------- 252

Qy         197 KDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEF-EPSNSHMVKSYSLLF--- 252
                                    :: ::  |||   : | :   |  || |
Db         253 ----------------------------SYSTVGSSSNTVADESLNQKPKPSSLRFDEP 283

Qy         253 -RVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLL 311
                 || |                 |  |  |:|: :|:         ::  |  | : :|
Db         284 HSKQRPAR---------------VPSEKFAKRRGSRD---------ISPADMYRSIMML 318

Qy         312 DGEYEVAMQEMEECPI---------------SKKRATWETILDGKRLPPFETFSQGPTL 355
               |::| |  | |  |:                  |:||    ::| ::   :  ::  |
Db         319 QGKHE-ATAEDEGPPLVIGSPADGTRYKSNVNKLKKATG---INGNKIKIRDKGNESNTD 374

Qy         356 QFTLRWTGETNDKSTAPIAKPLATRNSESLHQE--NKPGSVKPTQT-------IAVKESL 406
               | ::  : | |  ||| ::      :  |||    ||  |:: | |        ||: ::
Db         375 QNSV--SSEAN--STASVS------DESSLHTNFGNKVPSLR-TNTHRSNSGPIAITDNA 423

Qy         407 TTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTR---QQTEARDDLHCPWCTLNCRKLYS- 462
                ||     :|  |:          ||       |   ::    |||: |  ::  ||  |
Db         424 ETD-----KKHGPS---------IQFDITKPPRKISKRVSTFDDLN-PKSSVLYRKKASK 468

Qy         463 --LLKHLKLCHSRFIFNYVYHPKGARIDVSI-NECYDGSYAGNPQDIHRQPGFAFSRNGP 519
                 |:||                ||  ||   |       ||   |              | |
Db         469 KYLMKHF-------------PKARRIRQQIKRRLSTGSIEKN-----------SSNNV 502

Qy         520 VKRTPITHILVCRPKRTKASMSE--------------FLESEDGEVEQQRTYSSGHNRL 564
               | |||              | :                | ::| |: | :|   |
Db         503 SDRKPIT-----------DMDDDDDDDDNDGDNNEEYFADNESGD-EDERVQQSEP--- 546

Qy         565 YFHSDTCLPLRPQEME---------------VDSEDEKDPEWLREKTITQIEEFSDVNE 608
                 |||: |    |: |                |    :     | :||  : |  |:||
Db         547 --HSDSELKSHQQQQEKHQLQQNLHRMYKTKSFDDNRSRAVPMERSRTI-DMAEAKDLNE 603

Qy         609 ------GEKEVMKLWNLH---------------VMKHG-FIADNQMNHACMLFVENYGQK 646
                     :| | : |  |                : :|| : :|:  |:   |  | |
Db         604 LARTPDFQKMVYQNWKAHHRKKPNFRKRGWNNKIFEHGPYASDSDRNYPDN---SNTG-- 658

Qy         647 IIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQN 706
                     | :||          ||:  | :         |    |::| :|| |
Db         659 --------NSILHYAE-------SILHHDGS--------HKNGSEEASSDSNENIYSTNG 695

Qy         707 GT----ANGFSEINSKEKA---------LETDSVSGVSKQSKKQKL 739
               |:     | :   |  |:          : |    :|| | |  |
Db         696 GSDHNGLNNYPTYNDDEEGYYGLHFDTDYDLDPRHDLSKGSGKTYL 741


RESULT 6
US-09-702-705-1812
; Sequence 1812, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes; Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1812
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-702-705-1812

  Query Match             3.7%;  Score 142.5;  DB 4;  Length 474;
  Best Local Similarity   39.0%;  Pred. No. 0.00097;
  Matches   41;  Conservative   10;  Mismatches   25;  Indels   29;  Gaps    5;

Qy           6 HGGGGGGGSGPSAGSGGGGFGGSA----------AVAAATASGGKS---------GGGS 45
               |||||||||  : | |||  || |           :  |  | || |         |||
Db         166 HGGGGGGGSSNAGGGGGGASGGGANSKPAQKKSCGSKVAGGAGGGVSKPHAKLILAGGG- 224

Qy          46 CGGGGSYSASSSSSAAAAAG-AAVLPVKKPKMEHVQADHELFLQA 89
                 ||| : :|:::| ||  || ||:||:         |||    :|
Db         225 -GGGKAAAAAAASFAAEQAGAAALLPL------GAAADHHSLYKA 262


RESULT 7
US-09-736-457-1812
; Sequence 1812, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1812
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-736-457-1812

  Query Match             3.7%;  Score 142.5;  DB 4;  Length 474;
  Best Local Similarity   39.0%;  Pred. No. 0.00097;
  Matches   41;  Conservative   10;  Mismatches   25;  Indels   29;  Gaps    5;

Qy           6 HGGGGGGGSGPSAGSGGGGFGGSA----------AVAAATASGGKS---------GGGS 45
               |||||||||  : | |||  || |           :  |  | || |         |||
Db         166 HGGGGGGGSSNAGGGGGGASGGGANSKPAQKKSCGSKVAGGAGGGVSKPHAKLILAGGG- 224

Qy          46 CGGGGSYSASSSSSAAAAAG-AAVLPVKKPKMEHVQADHELFLQA 89
                 ||| : :|:::| ||  || ||:||:         |||    :|
Db         225 -GGGKAAAAAAASFAAEQAGAAALLPL------GAAADHHSLYKA 262


RESULT 8
US-09-059-584-51
; Sequence 51, Application US/09059584
; Patent No. 6440701
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Myers, Lisa E
;    APPLICANT:  Schryvers, Anthony B
;    APPLICANT:  Harkness, Robin E
;    APPLICANT:  Loosmore, Sheena M.
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Yang, Yan-Ping
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Transferrin Receptor Genes of Moraxella
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,584
;      FILING DATE:  14-APR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/778,570
;      FILING DATE:  03-JAN-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24973
;      REFERENCE/DOCKET NUMBER:  1038-794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  51:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  712 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-059-584-51

  Query Match             3.7%;  Score 142.5;  DB 4;  Length 712;
  Best Local Similarity   19.8%;  Pred. No. 0.0018;
  Matches  100;  Conservative   76;  Mismatches  169;  Indels  161;  Gaps   20;

Qy          9 GGGGGSGP----------SAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSS 58
              || ||| |          ||: | | ||: :   | |: | :|| | | |   |||:
Db         22 GGSGGSNPPAPTPIPNAGGAGNAGSGTGGAGSTDNA-ANAGSTGGASSGTG---SAST-- 75

Qy         59 SAAAAAGAAVLPVKKPKMEHVQADHEL--FLQAFEKPTQIYRF-LRTRNLIAPIFLHRTL 115
                        :||| : |  |      :  ::|   |   |: || |
Db         76 ------------QKPKYQDVPTDKNKKDEVSGIQEPAMGYGVELKLRNWIPQ------- 115

Qy        116 TYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNS 175
                  |    ||              | |::|: : :                   |  ||
Db        116 EQEEHAKINTN------------DVVKLEGDLKHN------------------PFDNS 143

Qy        176 --ENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVS 233
                :| :||  ::     | :::::::   |::     :  |   : :          |||
Db        144 IWQNIKNSKEVQT----VYNQEKQNIENQIKKENKELDKTALKALIEKVLDDYLTSLAKP 199

Qy        234 SNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEK 293
                |   ::||  :: :     ||  :   :| |      |||: :::
Db        200 IYEKNINDSHDKQNKAR----TRDLKYVRSGYIYRSGYSNIDIQKKI------------ 242

Qy        294 TFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPIS--KKRATWETILDGKRLPPFETFSQ 351
                  | ||           :    |  :: |:|  | : ||: : | |:   | :| :
Db        243 ----AKTGFD--------GALFYKGTQTAKQLPVSEVKYKGTWDFMTDAKKGQSFSSFER 290

Qy        352 GPTLQFTLRWTGE-----TNDKS--------------TAPIAKPLATRNSESLHQENKP 391
                  ::   : |     |:||:                |   :|   |    |  |::
Db        291 RAGDRYSAMSSHEYPSLLTDDKNKPDNYNDEYGHSSEFTVDFSKKSLTGGLFSNLQDHHK 350

Qy        392 GSVKPTQTIAV----------------------KESLTTDLQTRKEKDTPNENRQKLR 427
                | |  |:   :                    |   |:|   | |      | ::|
Db        351 GKVTKTKRYDINARIHGNRFRGSATAINKDNESKAKHPFTSDADNRLEGGFYGPNAEELA 410

Qy        428 IFYQFLYNNN-------TRQQTEARD 446
                 :|| ::|        :|::||::
Db        411 --GKFLTDDNKLFGVFGAKQESEAKE 434


RESULT 9
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
;  GENERAL INFORMATION:
;    APPLICANT:  Sivaraja, Mohanram
;    APPLICANT:  Strulovici, Berta
;    APPLICANT:  Flores, Osvaldo A.
;    TITLE OF INVENTION:  High Throughput In Vitro Screening Assay
;    TITLE OF INVENTION:  for Transcription Modulators
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/052,995
;      FILING DATE:  31-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parent, Annette S.
;      REGISTRATION NUMBER:  42,058
;      REFERENCE/DOCKET NUMBER:  018781-000600US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-052-995-1

  Query Match             3.6%;  Score 141;  DB 3;  Length 201;
  Best Local Similarity   44.1%;  Pred. No. 0.00035;
  Matches   26;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAG 65
              ||||||| ||  | |||| ||           ||  |||  ||||                |
Db         92 GGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150


-RESULT 10
US-09-053-003-40
; Sequence 40, Application US/09053003
; Patent No. 6207391
;  GENERAL INFORMATION:
;    APPLICANT:  Wu, Pengguang
;    APPLICANT:  McKinney, Judi
;    TITLE OF INVENTION:  High-Throughput Screening Assays for
;    TITLE OF INVENTION:  Modulators of STAT4 and STAT6 Activity
```

```
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/053,003
;      FILING DATE:  31-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Timothy L.
;      REGISTRATION NUMBER:  35,367
;      REFERENCE/DOCKET NUMBER:  018781-000800US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1..97
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 1-97 may be
;      OTHER INFORMATION:  present or absent"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  105..201
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 105-201 may be
;      OTHER INFORMATION:  present or absent"
US-09-053-003-40

  Query Match             3.6%;  Score 141;  DB 3;  Length 201;
  Best Local Similarity   44.1%;  Pred. No. 0.00035;
  Matches   26;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAG 65
              ||||||| ||  | |||| ||          ||  |||  ||||               |
Db         92 GGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150


RESULT 11
US-09-054-281-22
; Sequence 22, Application US/09054281
; Patent No. 6444421
;  GENERAL INFORMATION:
;    APPLICANT:  Chung, Jay H.
;    TITLE OF INVENTION:  Methods for Detecting Intermolecular
;    TITLE OF INVENTION:  Interactions In Vivo and In Vitro
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/054,281
;      FILING DATE:  02-APR-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/080,234
;      FILING DATE:  03-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/826,622
;      FILING DATE:  03-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Timothy L.
;      REGISTRATION NUMBER:  35,367
;      REFERENCE/DOCKET NUMBER:  015280-295200US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1..97
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 1-97 may be
;      OTHER INFORMATION:  present or absent"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  105..201
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 105-201 may be
;      OTHER INFORMATION:  present or absent"
US-09-054-281-22

  Query Match             3.6%;  Score 141;  DB 4;  Length 201;
  Best Local Similarity   44.1%;  Pred. No. 0.00035;
  Matches   26;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAG 65
              ||||||| ||  | |||| ||          ||  |||  ||||               |
Db         92 GGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150


RESULT 12
US-09-478-948-6
; Sequence 6, Application US/09478948
; Patent No. 6465258
; GENERAL INFORMATION:
;  APPLICANT: Shan, Bei
;  APPLICANT:  Okamoto, Arthur Y.
;  APPLICANT:  Tularik Inc.
;  TITLE OF INVENTION: FXR Receptor-Mediated Modulation of Cholesterol
;  TITLE OF INVENTION:  Metabolism
;  FILE REFERENCE: 018781-001310US
;  CURRENT APPLICATION NUMBER: US/09/478,948
;  CURRENT FILING DATE:  2000-01-06
;  PRIOR APPLICATION NUMBER: US 60/115,249
;  PRIOR FILING DATE: 1999-01-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 201
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:flexible linker
;   NAME/KEY: MOD_RES
```

```
;   LOCATION: (1)..(97)
;   OTHER INFORMATION: Gly at positions 1-97 may be present or absent
;   NAME/KEY: MOD_RES
;   LOCATION: (105)..(201)
;   OTHER INFORMATION: Gly at positions 105-201 may be present or absent
US-09-478-948-6

  Query Match             3.6%;  Score 141;  DB 4;  Length 201;
  Best Local Similarity   44.1%;  Pred. No. 0.00035;
  Matches   26;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          7 GGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAG 65
              ||||||| ||  | |||| ||          ||  |||  ||||              |
Db         92 GGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150


RESULT 13
US-07-789-915A-8
; Sequence 8, Application US/07789915A
; Patent No. 5212058
;  GENERAL INFORMATION:
;    APPLICANT:  Baker, Rohan T.
;    APPLICANT:  Tobias, John W.
;    APPLICANT:  Varshavsky, Alexander
;    TITLE OF INVENTION:  Ubiquitin-Specific Proteases
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/789,915A
;      FILING DATE:  19911108
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  MIT-5091AA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  912 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-789-915A-8

  Query Match             3.6%;  Score 139.5;  DB 1;  Length 912;
  Best Local Similarity   18.1%;  Pred. No. 0.0048;
  Matches  141;  Conservative  107;  Mismatches  264;  Indels  269;  Gaps   31;

Qy         28 SAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHELFL 87
              |  :     | |  |  | | | ||: | :
Db        102 SNGITNNNGSSGNQGANSSGSGMSYNKSHT---------------------------- 131

Qy         88 QAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGE 147
                        |    : | |          |:  || :|          : ::|
Db        132 ---------YHHNYSNNHI---------PMMASPNSGSN-----------AGMKKQTNS 161

Qy        148 QESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVP 207
                 :  ||              || :| |  :       | |    | |:    :::
Db        162 SNGNGSSA------------TSPSYSSYNSSSQYD----LYKFDVTKLKN----LKENS 200

Qy        208 TGKKQVPLIPDLNQTKPG-----------NFPSLAVSSNEFEPSN------SHMVKSYSL 250
              :   |:||   :| |:             |  :| ::|   | |:       |  ||:|:
Db        201 SNLIQLPLF--INTTEAEFAAASVQRYELNMKALNLNSESLENSSVEKSSAHHHTKSHSI 258

Qy        251 LFRVTRPGRREFNGMINGETN-ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQ 309
                        : |  :  ||: |  | ::: |    |    |  :|| | :
Db        259 ---------PKHNEEVKTETHGEEEDAHDKKPHASKDAHELKKKTEVKK----------- 298

Qy        310 LLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKS 369
                  | | |:  |  | : :||   ::| |   | |:  :          | :|:  |
Db        299 ------EDAKQDRNEKVIQEPQATVLPVVDKKE--PEESVEEN---------TSKTSSPS 341

Qy        370 TAPIA----KPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQK 425
               :| |      :|:  :|    ||  |        :  :       |      |  ::
Db        342 PSPPAAKSWSAIASDAIKSRQASNKTVSGSMVTKTPISGTTAGVSSTNMAAATIGKSSSP 401

Qy        426 LRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGA 485
              |             :| : :|  : |  |      | |:   |::|       :||   |
Db        402 L-----------LSKQPQKKDKKYVPPSTKGIEPLGSIA--LRMCFDPDFISYVLRNKDV 448

Qy        486 RIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRP----------KR 535
                : ::          |: |  :  |         : :  :| |:|          :
Db        449 ENKIPVHSII-------PRGIINRANICF------MSSVLQVLLYCKPFIDVINVLSTRN 495

Qy        536 TKASM-----------------------SEFLESEDG--EVEQQRTYSSGHNRLYFHSD 569
              | : :                       :|||: |   :  :     |  :: : |
Db        496 TNSRVGTSSCKLLDACLTMYKQFDKETYEKKFLENADDAEKTTESDAKKSSKSKSFQHCA 555

Qy        570 TCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKH----- 624
              |   ::| |                 ||:: | :| |:  | :|  : :  |::
Db        556 TADAVKPDEF--------------YKTLSTIPKFKDLQWGHQEDAEEFLTHLLDQLHEEL 601

Qy        625 ----GFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHLVSMH---DFNLISIMSIDKA 677
                    : ||:          ::|  | |  : | : | :  :| :   :|        | |
Db        602 ISAIDGLTDNE--------IQNMLQSINDEQL-KVFFIRNLSRYGKAEF-------IKNA 645

Qy        678 VTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQ 737
                :|:|: :|         |:: |||     ||: |            ||| ||: ||
Db        646 SPRLKELIEKY------GVINDDSTEE-----NGWHE-----------VSGSSKRGKKT 682

Qy        738 K 738
              |
Db        683 K 683


RESULT 14
US-08-005-002C-8
; Sequence 8, Application US/08005002C
; Patent No. 5494818
;  GENERAL INFORMATION:
;    APPLICANT:  Baker, Rohan T.
;    APPLICANT:  Tobias, John W.
;    APPLICANT:  Varshavsky, Alexander
;    TITLE OF INVENTION:  Ubiquitin-Specific Proteases
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  Maine
;      COUNTRY:  U.S.A.
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/005,002C
;      FILING DATE:  15-JAN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/789,915
```

```
;      FILING DATE:  08-NOV-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  MIT-5091AAZ
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  207-363-0558
;      TELEFAX:  207-363-0528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  912 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-005-002C-8

  Query Match             3.6%;  Score 139.5;  DB 1;  Length 912;
  Best Local Similarity   18.1%;  Pred. No. 0.0048;
  Matches  141;  Conservative  107;  Mismatches  264;  Indels  269;  Gaps   31;

Qy         28 SAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHELFL 87
              |  :     | |  |  | | | ||: | :
Db        102 SNGITNNNGSSGNQGANSSGSGMSYNKSHT---------------------------- 131

Qy         88 QAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGE 147
                       |    : | |           |:  || :|             : ::|
Db        132 ---------YHHNYSNNHI---------PMMASPNSGSN-----------AGMKKQTNS 161

Qy        148 QESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVP 207
               :  ||              || :| |  :       | |    | |:    :::
Db        162 SNGNGSSA------------TSPSYSSYNSSSQYD----LYKFDVTKLKN----LKENS 200

Qy        208 TGKKQVPLIPDLNQTKPG-----------NFPSLAVSSNEFEPSN------SHMVKSYSL 250
              :   |:||   :| |:              |  :| ::|   | |:       |  ||:|:
Db        201 SNLIQLPLF--INTTEAEFAAASVQRYELNMKALNLNSESLENSSVEKSSAHHHTKSHSI 258

Qy        251 LFRVTRPGRREFNGMINGETN-ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQ 309
                     : |  :  ||: |  | ::: |   |   |  :|| | :
Db        259 ---------PKHNEEVKTETHGEEEDAHDKKPHASKDAHELKKKTEVKK----------- 298

Qy        310 LLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKS 369
                    | |:  |  | : :||   ::| |   | |:  :           | :|:  |
Db        299 ------EDAKQDRNEKVIQEPQATVLPVVDKKE--PEESVEEN---------TSKTSSPS 341

Qy        370 TAPIA----KPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQK 425
               :| |      :|:   :|    ||  |         :  :       |      |  ::
Db        342 PSPPAAKSWSAIASDAIKSRQASNKTVSGSMVTKTPISGTTAGVSSTNMAAATIGKSSSP 401

Qy        426 LRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGA 485
              |             :| : :|  : |  |      | |:   |::|       :||   |
Db        402 L-----------LSKQPQKKDKKYVPPSTKGIEPLGSIA--LRMCFDPDFISYVLRNKDV 448

Qy        486 RIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRP----------KR 535
                : ::          |: |  :    |        : :  :| |:|          :
Db        449 ENKIPVHSII-------PRGIINRANICF------MSSVLQVLLYCKPFIDVINVLSTRN 495

Qy        536 TKASM-----------------------SEFLESEDG--EVEQQRTYSSGHNRLYFHSD 569
              | : :                       :|||: |   :  :      |  :: : |
Db        496 TNSRVGTSSCKLLDACLTMYKQFDKETYEKKFLENADDAEKTTESDAKKSSKSKSFQHCA 555

Qy        570 TCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKH----- 624
              |   ::| |                 ||:: | :| |:  | :|  : :  |::
Db        556 TADAVKPDEF-------------YKTLSTIPKFKDLQWGHQEDAEEFLTHLLDQLHEEL 601

Qy        625 ----GFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHLVSMH---DFNLISIMSIDKA 677
                  : ||:        ::|  | |  : | : | :  :| :   :|         | |
Db        602 ISAIDGLTDNE--------IQNMLQSINDEQL-KVFFIRNLSRYGKAEF-------IKNA 645

Qy        678 VTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQ 737
               :|:|: :|          |:: |||     ||: |            ||| ||: ||
Db        646 SPRLKELIEKY------GVINDDSTEE-----NGWHE-----------VSGSSKRGKKT 682

Qy        738 K 738
```

```
              |
Db        683 K 683


RESULT 15
US-08-487-203A-8
; Sequence 8, Application US/08487203A
; Patent No. 5683904
;  GENERAL INFORMATION:
;    APPLICANT:  Baker, Rohan T.
;    APPLICANT:  Tobias, John W.
;    APPLICANT:  Varshavsky, Alexander
;    TITLE OF INVENTION:  Ubiquitin-Specific Proteases
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  Maine
;      COUNTRY:  U.S.A.
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,203A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/005,002
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  MIT-5091A3Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  207-363-0558
;      TELEFAX:  207-363-0528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  912 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-487-203A-8

  Query Match             3.6%;  Score 139.5;  DB 1;  Length 912;
  Best Local Similarity   18.1%;  Pred. No. 0.0048;
  Matches  141;  Conservative  107;  Mismatches  264;  Indels  269;  Gaps   31;

Qy         28 SAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHELFL 87
              |  :     | |  |  | | | ||: | :
Db        102 SNGITNNNGSSGNQGANSSGSGMSYNKSHT---------------------------- 131

Qy         88 QAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGE 147
                       |    : | |           |:  || :|             : ::|
Db        132 ---------YHHNYSNNHI---------PMMASPNSGSN-----------AGMKKQTNS 161

Qy        148 QESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVP 207
               :  ||              || :| |  :       | |    | |:    :::
Db        162 SNGNGSSA------------TSPSYSSYNSSSQYD----LYKFDVTKLKN----LKENS 200

Qy        208 TGKKQVPLIPDLNQTKPG-----------NFPSLAVSSNEFEPSN------SHMVKSYSL 250
              :   |:||   :| |:              |  :| ::|   | |:       |  ||:|:
Db        201 SNLIQLPLF--INTTEAEFAAASVQRYELNMKALNLNSESLENSSVEKSSAHHHTKSHSI 258

Qy        251 LFRVTRPGRREFNGMINGETN-ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQ 309
                     : |  :  ||: |  | ::: |   |   |  :|| | :
Db        259 ---------PKHNEEVKTETHGEEEDAHDKKPHASKDAHELKKKTEVKK---------- 298

Qy        310 LLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKS 369
```

```
                 | | |:  |  | : :|| . ::| |   | |:  :          | :|:  |
Db        299 ------EDAKQDRNEKVIQEPQATVLPVVDKKE--PEESVEEN---------TSKTSSPS 341

Qy        370 TAPIA----KPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQK 425
               :| |      :|:   :|    ||  |         :  :       |      |  ::
Db        342 PSPPAAKSWSAIASDAIKSRQASNKTVSGSMVTKTPISGTTAGVSSTNMAAATIGKSSSP 401

Qy        426 LRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGA 485
               |           :| : :|  : |  |     | |:   |::|      :||   |
Db        402 L-----------LSKQPQKKDKKYVPPSTKGIEPLGSIA--LRMCFDPDFISYVLRNKDV 448

Qy        486 RIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRP----------KR 535
               : ::          |: |  :    |        : :  :| |:|          :
Db        449 ENKIPVHSII-------PRGIINRANICF------MSSVLQVLLYCKPFIDVINVLSTRN 495

Qy        536 TKASM-----------------------SEFLESEDG--EVEQQRTYSSGHNRLYFHSD 569
               | : :                        :|||: |   :  :     |  :: : |
Db        496 TNSRVGTSSCKLLDACLTMYKQFDKETYEKKFLENADDAEKTTESDAKKSSKSKSFQHCA 555

Qy        570 TCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKH----- 624
               |   ::| |                ||:: | :| |:  | :|  : :  |::
Db        556 TADAVKPDEF-------------YKTLSTIPKFKDLQWGHQEDAEEFLTHLLDQLHEEL 601

Qy        625 ----GFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHLVSMH---DFNLISIMSIDKA 677
                   : ||:        ::|  | |  : | : |  :| :   :|       |  |
Db        602 ISAIDGLTDNE--------IQNMLQSINDEQL-KVFFIRNLSRYGKAEF-------IKNA 645

Qy        678 VTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQ 737
               :|:|: :|          |:: |||     ||: |            ||| ||: ||
Db        646 SPRLKELIEKY------GVINDDSTEE-----NGWHE-----------VSGSSKRGKKT 682

Qy        738 K 738
               |
Db        683 K 683
```

Search completed: September 2, 2003, 11:06:56
Job time : 26.3289 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37 ; Search time 54.1833 Seconds
(without alignments)
3519.551 Million cell updates/sec

Title: US-09-874-162A-5
Perfect score: 3885
Sequence: 1 MAPQKHGGGGGGGSGPSAGS.....KALETDSVSGVSKQSKKQKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3885	100.0	803	4	Q15022	Q15022 homo sapien
2	3878	99.8	739	4	Q96BD9	Q96bd9 homo sapien
3	1177	30.3	855	5	Q9VW55	Q9vw55 drosophila
4	1177	30.3	900	5	Q9NJG9	Q9njg9 drosophila
5	550	14.2	114	11	Q99L07	Q99l07 mus musculu
6	238.5	6.1	626	10	Q8L6Y4	Q8l6y4 arabidopsis
7	236	6.1	631	10	Q93V59	Q93v59 arabidopsis
8	221	5.7	445	10	Q8W5B2	Q8w5b2 arabidopsis
9	217	5.6	440	10	Q94CF5	Q94cf5 arabidopsis
10	190	4.9	692	10	Q9ZNT9	Q9znt9 arabidopsis
11	180	4.6	632	10	Q9ZQP0	Q9zqp0 arabidopsis
12	171	4.4	570	13	Q8AWB3	Q8awb3 lampetra fl
13	169.5	4.4	1745	10	Q9MAK1	Q9mak1 arabidopsis
14	166.5	4.3	620	5	O62007	O62007 branchiosto
15	165	4.2	2359	5	Q8I519	Q8i519 plasmodium
16	164	4.2	478	13	Q98TQ7	Q98tq7 lampetra fl
17	162.5	4.2	1390	4	Q9ULU8	Q9ulu8 homo sapien
18	162	4.2	1274	4	Q8NFR0	Q8nfr0 homo sapien
19	161	4.1	342	5	Q9VKR8	Q9vkr8 drosophila
20	158.5	4.1	730	10	Q9LP25	Q9lp25 arabidopsis
21	158	4.1	1103	5	Q9VY72	Q9vy72 drosophila
22	156.5	4.0	1011	2	Q9AHL0	Q9ahl0 borrelia bu
23	156	4.0	620	5	O62004	O62004 branchiosto
24	155	4.0	987	10	Q8LIX8	Q8lix8 oryza sativ
25	154.5	4.0	1081	11	Q8C7Q2	Q8c7q2 mus musculu
26	154.5	4.0	1289	11	Q62717	Q62717 rattus norv
27	154	4.0	629	13	Q90WR5	Q90wr5 lampetra fl
28	153	3.9	433	13	Q910C2	Q910c2 rana catesb
29	153	3.9	558	13	Q9PUB5	Q9pub5 brachydanio
30	153	3.9	738	5	O02402	O02402 pinctada fu
31	153	3.9	1777	5	Q8I3P4	Q8i3p4 plasmodium
32	152	3.9	427	5	Q95TA5	Q95ta5 drosophila
33	152	3.9	1422	10	Q9ZUR3	Q9zur3 arabidopsis
34	151.5	3.9	393	5	Q18880	Q18880 caenorhabdi
35	151	3.9	2165	4	Q8NFD5	Q8nfd5 homo sapien
36	150.5	3.9	618	11	Q8CIS9	Q8cis9 rattus norv
37	150	3.9	1957	4	Q8IZY8	Q8izy8 homo sapien
38	149.5	3.8	2639	5	O76786	O76786 antheraea p
39	149	3.8	284	4	Q9H524	Q9h524 homo sapien
40	149	3.8	512	4	Q9Y6P1	Q9y6p1 homo sapien
41	149	3.8	706	4	Q9UI36	Q9ui36 homo sapien
42	149	3.8	1189	5	Q8I825	Q8i825 dictyosteli
43	148.5	3.8	1163	5	Q8IKP5	Q8ikp5 plasmodium
44	148.5	3.8	1611	10	Q9SU69	Q9su69 arabidopsis
45	148	3.8	847	4	Q8IXI9	Q8ixi9 homo sapien

ALIGNMENTS

RESULT 1
Q15022
ID Q15022 PRELIMINARY; PRT; 803 AA.
AC Q15022;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA0160 (Fragment).
GN KIAA0160.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
DR EMBL; D63881; BAA09931.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; ZnF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 803 AA; 89963 MW; CDFB901A35F29A7C CRC64;

Query Match 100.0%; Score 3885; DB 4; Length 803;
Best Local Similarity 100.0%; Pred. No. 1.4e-263;
Matches 739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 124

Qy         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        125 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 184

Qy        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 244

Qy        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 304

Qy        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 364

Qy        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 424

Qy        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 484

Qy        421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 544

Qy        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 604

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 664

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 724

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 784

Qy        721 ALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||
Db        785 ALETDSVSGVSKQSKKQKL 803
```

```
RESULT 2
Q96BD9
ID   Q96BD9      PRELIMINARY;      PRT;   739 AA.
AC   Q96BD9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Joined to JAZF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC015704; AAH15704.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ   SEQUENCE   739 AA;  83054 MW;  A8830EBC3FD38D56 CRC64;

  Query Match           99.8%;  Score 3878;  DB 4;  Length 739;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-263;
  Matches  738;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60

Qy         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120

Qy        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180

Qy        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLNPDLNQTKPGNFPSLAVSSNEFEPS 240

Qy        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300

Qy        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360

Qy        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420

Qy        421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480

Qy        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720

Qy        721 ALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||
Db        721 ALETDSVSGVSKQSKKQKL 739
```

```
RESULT 3
Q9VW55
ID   Q9VW55      PRELIMINARY;      PRT;   855 AA.
AC   Q9VW55; Q8T9D8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG8013 protein (SD04959p).
GN   SU(Z)12 OR CG8013.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.

RN [6]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003515; AAF49094.2; -.
DR EMBL; AY069809; AAL39954.1; -.
DR FlyBase; FBgn0020887; Su(z)12.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ SEQUENCE 855 AA; 95316 MW; C10FFC6013954105 CRC64;

Query Match 30.3%; Score 1177; DB 5; Length 855;
Best Local Similarity 34.8%; Pred. No. 6.8e-74;
Matches 281; Conservative 136; Mismatches 225; Indels 166; Gaps 26;

```
Qy         43 GGSCGG--GGSYSASSSSSAAAAAGAAVLPVK--KPKMEHVQADHELFLQAFEKPTQIYR 98
              |  :  |  | :: |  :|:    ||: | |    : |:   | : ||||||||||||||
Db         15 GSAANGIIGLTHGAPDASN----AGSTVPPTAEGQVKLNGHQQEQELFLQAFEKPTQIYR 70

Qy         99 FLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSA-HL 157
              :|| |:   ||||:|||:||  | || | || :|:|: || :      |:| ::|  :|
Db         71 YLRNRHETNPIFLNRTLSYMKERMSRNNKKRISFQVNSMLESI-----TQKSEAVSQNYL 125

Qy        158 QLTFTGFFHKNDKPSPNSENEQ----------NSVTLEVLLVKVCHKKRKDVSCPIRQV 206
              : :     |       |   |            ||::|  | |:  ||||  :  :::
Db        126 HVIYDSLHEKLPARLDNESGEDLLQEQLLCEAGESVSVETTLYKITRSKRKDSTLDFQEL 185

Qy        207 PTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEP-SNSHMVKSYSLLFRVTRPGRREFNGM 265
              :  |:  |  : : |  ::::     |   | :  | ||||:         :
Db        186 LSKCSQIVYNP---KDRVGEHATISIPLQTMRPMGEQHTL--YKLLFRIK---------V 231

Qy        266 INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEEC 325
              ::  |    | | | |  ||:| : || | ::: :::|:    : :||||  :| :
Db        232 LSPSTCN--DENAETPP-NKRSRPN-EKMFGSELILYEKSSGF-ITEGEYEAMLQPLNST 286

Qy        326 PI---SKKRATWETILDGKRLP---PFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLAT 379
              |   | |: ||||: |   :|    :: : | | |:| |     :|::    |: |
Db        287 SIKSFSPKKCTWETMPD-SYIPLSLTYDVYQQSPMLKFHLTL---SNEQLPEMISAPELQ 342

Qy        380 R------------NSESLHQEN-----KPGSVKPTQTIAVKESLTTDLQTRKEKDTPNE 421
              |            |: : :  |     | ||     |:                | ||
Db        343 RYVQHLDAVAEMNYNNNNYNNNNCSGLKNGSGGGNSTVC-------------KTTP-- 386

Qy        422 NRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYH 481
              : ::| | |:|:|||||||    :|:|||| |:| :||:|||||||||:|| | |
Db        387 --EHIQIVYNFMYSNNTRQQTEYTQELNCPWCGLDCLRLYALLKHLKLCHARFNFTYQPA 444

Qy        482 PKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRN-GPVKRTPITHILVCRPKRTKASM 540
              ||||||:||: |||||||:| |:    | :|:|  |||:|| :| ::||||:| |  :
Db        445 GSGARIDVTINDAYDGSYAGSPYDLAGPSGSSFARTCGPVRRTSVTSLMVCRPRRQKTCL 504

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              |||| :: |:  ||:| :|||||| |::|||: |:|:::||| | || |||:|||  |
Db        505 DEFLELDEDEISNQRSYITGHNRLYHHTETCLPVHPKELDIDSEGESDPLWLRQKTIQMI 564

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              :||||||||||:|||||||||:|||: | |:  || :|::  | :|::||| |||:||:
Db        565 DEFSDVNEGEKELMKLWNLHVMRHGFVGDCQLPIACEMFLDAKGTEIVRKNLYRNFILHM 624

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGES-------------------------- 693
              |: |: ||:  :: | | ||: :  |   |:
Db        625 CSLFDYGLIAAETVYKTVQKLQGLLSKYAAGQELMQRQREEQLKYWLDVGMHKKQEDPKT 684

Qy        694 -------ASPANEEITE---------------------------------EQNGTANG 711
                    | ||::  |                                    :  || ||
Db        685 LKSPQKPAPPADQASTSSASTSGSGSGSSSMQPPKRMPAHLKRGSAASSPGVQSKGTENG 744

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
              :  ||      : :   |:|:|  | |
```

```
Db        745 TNGSNS-----SSSNSKNVAKKSADQPL 767


RESULT 4
Q9NJG9
ID   Q9NJG9      PRELIMINARY;      PRT;   900 AA.
AC   Q9NJG9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Su(z)12.
GN   SU(Z)12 OR CG8013.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Birve A., Rasmuson-Lestander A., Larsson J.;
RT   "Suppressor of zeste 12.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF149047; AAF73149.1; -.
DR   FlyBase; FBgn0020887; Su(z)12.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ   SEQUENCE   900 AA;  100103 MW;  53BA0D83C49EC92F CRC64;

  Query Match             30.3%;  Score 1177;  DB 5;  Length 900;
  Best Local Similarity   34.8%;  Pred. No. 7.2e-74;
  Matches  281;  Conservative  136;  Mismatches  225;  Indels  166;  Gaps   26;

Qy         43 GGSCGG--GGSYSASSSSSAAAAAGAAVLPVK--KPKMEHVQADHELFLQAFEKPTQIYR 98
              |  :  |  | :: |  :|:    ||: | |    : |:   | : ||||||||||||||
Db         15 GSAANGIIGLTHGAPDASN----AGSTVPPTAEGQVKLNGHQQEQELFLQAFEKPTQIYR 70

Qy         99 FLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSA-HL 157
              :|| |:   ||||:|||:||  | || | || :|:|: ||  :      |:| ::|  :|
Db         71 YLRNRHETNPIFLNRTLSYMKERMSRNNKKRISFQVNSMLESI-----TQKSEAVSQNYL 125

Qy        158 QLTFTGFFHKNDKPSPNSENEQ----------NSVTLEVLLVKVCHKKRKDVSCPIRQV 206
               : :     |      |   |              ||::|  | |:   |||| :   :::
Db        126 HVIYDSLHEKLPARLDNESGEDLLQEQLLCEAGESVSVETTLYKITRSKRKDSTLDFQEL 185

Qy        207 PTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEP-SNSHMVKSYSLLFRVTRPGRREFNGM 265
              :   |:   |   : : |   ::::      |    | :  | ||||:           :
Db        186 LSKCSQIVYNP---KDRVGEHATISIPLQTMRPMGEQHTL--YKLLFRIK---------V 231

Qy        266 INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEEC 325
              ::  |    | | | |   ||:| : || | ::: :::|:    : :||||  :| :
Db        232 LSPSTCN--DENAETPP-NKRSRPN-EKMFGSELILYEKSSGF-ITEGEYEAMLQPLNST 286

Qy        326 PI---SKKRATWETILDGKRLP---PFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLAT 379
               |   | |: |||| |   :|    :: : | | |:| |     :|::    |: |
Db        287 SIKSFSPKKCTWETMPD-SYIPLSLTYDVYQQSPMLKFHLTL---SNEQLPEMISAPELQ 342

Qy        380 R------------NSESLHQEN-----KPGSVKPTQTIAVKESLTTDLQTRKEKDTPNE 421
              |            |: : :  |     | ||     |:                | ||
Db        343 RYVQHLDAVAEMNYNNNNYNNNNNCSGLKNGSGGGNSTVC-------------KTTP-- 386

Qy        422 NRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYH 481
               : ::| | |:|:|||||||   :|:|||| |:| :||:|||||||||:|| | |
Db        387 --EHIQIVYNFMYSNNTRQQTEYTQELNCPWCGLDCLRLYALLKHLKLCHARFNFTYQPA 444

Qy        482 PKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRN-GPVKRTPITHILVCRPKRTKASM 540
              ||||||:||: ||||||:| |:    | :|:|  |||:|| :| ::|||||| |  :
Db        445 GSGARIDVTINDAYDGSYAGSPYDLAGPSGSSFARTCGPVRRTSVTSLMVCRPRRQKTCL 504

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              |||| :: |:  ||:| :|||||| |::|||: |:|:::||| | || |||:|||  |
Db        505 DEFLELDEDEISNQRSYITGHNRLYHHTETCLPVHPKELDIDSEGESDPLWLRQKTIQMI 564

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              :|||||||||||:|||||||||:|||: | |:  || :|:::  | :|:::||| |||:||:
Db        565 DEFSDVNEGEKELMKLWNLHVMRHGFVGDCQLPIACEMFLDAKGTEIVRKNLYRNFILHM 624

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGES-------------------------- 693
               |: |: ||:  :: | | ||: :  |   |:
Db        625 CSLFDYGLIAAETVYKTVQKLQGLLSKYAAGQELMQRQREEQLKYWLDVGMHKKQEDPKT 684

Qy        694 -------ASPANEEITE----------------------------------EQNGTANG 711
                    | ||::  |                                      :  || ||
Db        685 LKSPQKPAPPADQASTSSASTSGSGSGSSSMQPPKRMPAHLKRGSAASSPGVQSKGTENG 744

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
               :  ||      : :   |:|:|  | |
Db        745 TNGSNS-----SSSNSKNVAKKSADQPL 767


RESULT 5
Q99L07
ID   Q99L07      PRELIMINARY;      PRT;   114 AA.
AC   Q99L07;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 12.8 kDa protein (Fragment).
GN   D11ERTD530E.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC003922; AAH03922.1; -.
DR   MGD; MGI:1261758; D11Ertd530e.
KW   Hypothetical protein.
FT   NON_TER      1      1
SQ   SEQUENCE   114 AA;  12849 MW;  A73CEF641A1E9228 CRC64;

  Query Match             14.2%;  Score 550;  DB 11;  Length 114;
  Best Local Similarity   94.7%;  Pred. No. 3.9e-31;
  Matches  108;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        626 FIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ 60

Qy        686 QKLEKGESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||:|:|||| |||||||||||| |||||||||| |||| ||||||||
Db         61 QKLEKGESATPSNEEIAEEQNGTANGFSETNSKEKALETDGVSGVPKQSKKQKL 114


RESULT 6
Q8L6Y4
ID   Q8L6Y4      PRELIMINARY;      PRT;   626 AA.
AC   Q8L6Y4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   AT5G51230:AT5G51240.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
```

```
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY140086; AAM98227.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   626 AA;  71033 MW;  E37A79F64DCF1D7A CRC64;

  Query Match             6.1%;  Score 238.5;  DB 10;  Length 626;
  Best Local Similarity   19.5%;  Pred. No. 2.4e-08;
  Matches  137;  Conservative  104;  Mismatches  258;  Indels  203;  Gaps   24;

Qy         77 EHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDD 136
Db         34 EEIAAEES--LAAYCKPVELYNIIQRRAIRNPLFLQRCLHYKIEAKHKRRIQMTVFL--- 88

Qy        137 MLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
Db         89 ------------SGAIDAGVQ------------------TQKLFPLYILLARLVSPK- 115

Qy        197 KDVSCPIRQ-----------VPTGKKQVP---------LIPDLN----QTKPGN----FP 228
Db        116 -----PVAEYSAVYRFSRACILTGGLGVDGVSQAQANFLLPDMNRLALEAKSGSLAILFI 170

Qy        229 SLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNR 288
Db        171 SFAGAQNSQFGIDSGKIHS----------------GNIGGHC-----LWSKIPLQ----- 204

Qy        289 EDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFET 348
Db        205 --------SLYASWQKSPNMDLGQRVDTVSLVEMQPCFIKLKSMSEEKCVS--------- 247

Qy        349 FSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTT 408
Db        248 -IQVPSNPLT----------SSSPQQVQVTISAEEVGSTEKSPYSSFSYNDIS-SSSLLQ 295

Qy        409 DLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLK 468
Db        296 IIRLR----TGN-------VVFNYRYYNNKLQKTEVTEDFSCPFCLVKCASFKGLRYHLP 344

Qy        469 LCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIH-RQPGFAF----------SR 516
Db        345 STHDLLNFEFWVTEEFQAVNVSLKTETMISKV-NEDDVDPKQQTFFFSRRRQKSQVRSSR 403

Qy        517 NGP--------VKRTPITHILVCRPKR------------------------------- 535
Db        404 QGPHLGLGCEVLDKTDDAHSVRSEKSRIPPGKHYERIGGAESGQRVPPGTSPADVQSCGD 463

Qy        536 -------TKASMSEFLESEDGEVEQQ--RTYSSGHNRLYFHSDTCLPLRPQEM--EVDSE 584
Db        464 PDYVQSIAGSTMLQFAKTRKISIERSDLRNRSLLQKRQFFHSHRAQPMALEQVLSDRDSE 523

Qy        585 DEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYG 644
Db        524 DEVDDDVADFEDRRMLDDFVDVTKDEKQMMHMWNSFVRKQRVLADGHIPWACEAFSRLHG 583

Qy        645 QKIIK-KNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ 685
Db        584 PIMVRTPHLIWCWRVFMVKLWNHGLLDARTMNNCNTFLEQLQ 625


RESULT 7
Q93V59
ID   Q93V59      PRELIMINARY;      PRT;   631 AA.
AC   Q93V59;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Embryonic flower 2.
GN   EMF2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yoshida N., Yanai Y.;
RT   "Arabidopsis, embryonic flower 2 gene.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yoshida N.;
RT   "embryonic flower 2.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB053262; BAB58957.1; -.
DR   EMBL; AB053171; BAB58956.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ   SEQUENCE   631 AA;  71680 MW;  F36FE92D0F62E610 CRC64;

  Query Match             6.1%;  Score 236;  DB 10;  Length 631;
  Best Local Similarity   19.4%;  Pred. No. 3.6e-08;
  Matches  137;  Conservative  104;  Mismatches  258;  Indels  208;  Gaps   24;

Qy         77 EHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDD 136
Db         34 EEIAAEES--LAAYCKPVELYNIIQRRAIRNPLFLQRCLHYKIEAKHKRRIQMTVFL--- 88

Qy        137 MLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
Db         89 ------------SGAIDAGVQ------------------TQKLFPLYILLARLVSPK- 115

Qy        197 KDVSCPIRQ-----------VPTGKKQVP---------LIPDLN----QTKPGN----FP 228
Db        116 -----PVAEYSAVYRFSRACILTGGLGVDGVSQAQANFLLPDMNRLALEAKSGSLAILFI 170

Qy        229 SLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNR 288
Db        171 SFAGAQNSQFGIDSGKIHS----------------GNIGGHC-----LWSKIPLQ----- 204

Qy        289 EDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFET 348
Db        205 --------SLYASWQKSPNMDLGQRVDTVSLVEMQPCFIKLKSMSEEKCVS--------- 247

Qy        349 FSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTT 408
Db        248 -IQVPSNPLT----------SSSPQQVQVTISAEEVGSTEKSPYSSFSYNDIS-SSSLLQ 295

Qy        409 DLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLK 468
Db        296 IIRLR----TGN-------VVFNYRYYNNKLQKTEVTEDFSCPFCLVKCASFKGLRYHLP 344

Qy        469 LCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIH-RQPGFAF---------- 514
Db        345 STHDLLNFEFWVTEEFQAVNVSLKTETMISKV-NEDDVDPKQQTFFFSSKKFRRRRQKSQ 403

Qy        515 ---SRNGP--------VKRTPITHILVCRPKR---------------------------- 535
Db        404 VRSSRQGPHLGLGCEVLDKTDDAHSVRSEKSRIPPGKHYERIGGAESGQRVPPGTSPADV 463

Qy        536 ------------TKASMSEFLESEDGEVEQQ--RTYSSGHNRLYFHSDTCLPLRPQEM-- 579
Db        464 QSCGDPDYVQSIAGSTMLQFAKTRKISIERSDLRNRSLLQKRQFFHSHRAQPMALEQVLS 523

Qy        580 EVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLF 639
Db        524 DRDSEDEVDDDVADFEDRRMLDDFVDVTKDEKQMMHMWNSFVRKQRVLADGHIPWACEAF 583

Qy        640 VENYGQKIIK-KNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ 685
Db        584 SRLHGPIMVRTPHLIWCWRVFMVKLWNHGLLDARTMNNCNTFLEQLQ 630
```

```
RESULT 8
Q8W5B2
ID   Q8W5B2      PRELIMINARY;      PRT;   445 AA.
AC   Q8W5B2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Vernalization 2 protein.
GN   VRN2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21575875; PubMed=11719192;
RA   Gendall A.R., Levy Y.Y., Wilson A., Dean C.;
RT   "The VERNALIZATION2 (VRN2) gene mediates the epigenetic regulation of
RT   vernalization in Arabidopsis.";
RL   Cell 107:525-535(2001).
DR   EMBL; AF284500; AAL32135.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ   SEQUENCE   445 AA;  51177 MW;  30A3391CE98D877C CRC64;

  Query Match             5.7%;  Score 221;  DB 10;  Length 445;
  Best Local Similarity   21.1%;  Pred. No. 2.6e-07;
  Matches   84;  Conservative   64;  Mismatches  151;  Indels  100;  Gaps   12;

Qy        413 RKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHS 472
              || |     :|    : : :   ||| |:|| |:|  ||:|:: |     |  ||   |
Db         58 RKRK-----SRSTGMVVFNYKDCNNTLQKTEVREDCSCPFCSMLCGSFKGLQFHLNSSHD 112

Qy        473 RFIFNYVYHPKGARIDVSI--------NECYDG------SYAGNPQDIHRQPGFAFSRNG 518
              | | :    :   ::||:         |  |       |    |:   || |    ||
Db        113 LFEFEFKLSEEYQTVNVSVKLNSFIFEEEGSDDDKFEPFSLCSKPRK-RRQRG---GRNN 168

Qy        519 PVKRTPITHILVCRPKRTKASMSEFLESEDG---------------------------- 549
                :|  :  : :  |  |  : :      ||.
Db        169 -TRRLKVCFLPLDSPSLTNGTENGITLLNDGNRGLGYPEATELAGQFEMTSNIPPAIAHS 227

Qy        550 ----------------------EVEQQRTYSSGH----NRLYFHSDTCLPLRPQEM--EV 581
                                  :: :|: :  |     | ::||    |:  :::  :
Db        228 SLDAGAKVILTSEAVVPATKTRKLSAERSEARSHLLLQKRQFYHSHRVQPMALEQVMSDR 287

Qy        582 DSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVE 641
              ||||| | :    :    :::| |||: ||: | |||  | |   |||  :: ||  |
Db        288 DSEDEVDDDVADFEDRQMLDDFVDVNKDEKQFMHLWNSFVRKQRVIADGHISWACEAFSR 347

Qy        642 NYGQKIIK-KNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ----------QKLE 689
               | ::: :  :|   : | |: : :  |:   :|:   | |   :           ::
Db        348 FYEKELHRYSSLFWCWRLFLIKLWNHGLVDSATINNCNTILENCRNSSDTTTTNNNSVD 407

Qy        690 KGESASPANEEITEEQNGTANGFSEINSKEKALETDSVS 728
               :   ::  |  | : |      :||:|     |: |
Db        408 RPSDSNTNNNNIVDHPN------DINNKNNVDNKDNNS 439


RESULT 9
Q94CF5
ID   Q94CF5      PRELIMINARY;      PRT;   440 AA.
AC   Q94CF5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 50.6 kDa protein.
GN   DL4450W.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA   Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA   Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA   Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA   Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene dl4450w.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA   Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA   Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA   Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA   Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Full Length cDNA of gene dl4450w.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY034902; AAK59409.1; -.
DR   EMBL; AY063047; AAL34221.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   440 AA;  50623 MW;  0E0AB2C55517BE6F CRC64;

  Query Match             5.6%;  Score 217;  DB 10;  Length 440;
  Best Local Similarity   21.2%;  Pred. No. 4.8e-07;
  Matches   80;  Conservative   62;  Mismatches  140;  Indels   96;  Gaps   11;

Qy        413 RKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHS 472
              || |     :|    : : :   ||| |:|| |:|  ||:|:: |     |  ||   |
Db         58 RKRK-----SRSTGMVVFNYKDCNNTLQRTEVREDCSCPFCSMLCGSFKGLQFHLNSSHD 112

Qy        473 RFIF-----------------NYVYHPKGARIDV----------------------SI 491
              | |                    ::::  :|:  |                      :
Db        113 LFEFEFKLLEEYQTVNVSVKLNSFIFEEEGSDDDKFEPFSLCSKPRKRRQRGGRNNTRRL 172

Qy        492 NECY-------------------DGSYA-GNPQDIHRQPGFAFSRNGPVKRTPITH--- 527
               |:                   ||:   | |:     |  : | |     | |
Db        173 KVCFLPLDSPSLANGTENGIALLNDGNRGLGYPEATELAGQFEMTSNIP---PAIAHSSL 229

Qy        528 -----ILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEM--E 580
                   :::       |: :  | :|  |    |::     | ::||    |:  :::  :
Db        230 DAGAKVILTTEAVVPATKTRKLSAERSEA---RSHLLLQKRQFYHSHRVQPMALEQVMSD 286

Qy        581 VDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFV 640
              |||||  | :    :    :::| |||: ||: | |||  | |   |||  :: || :|
Db        287 RDSEDEVDDDVADFEDRQMLDDFVDVNKDEKQFMHLWNSFVRKQRVIADGHISWACEVFS 346

Qy        641 ENYGQKI-IKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANE 699
               | :::    :|   : | |: : :  |:   :|:   | |   :            |
Db        347 RFYEKELHCYSSLFWCWRLFLIKLWNHGLVDSATINNCNTILENCR-----------NT 394

Qy        700 EITEEQNGTANGFSEINS 717
               :|   | : :  |: |:
Db        395 SVTNNNNNSVDHPSDSNT 412


RESULT 10
Q9ZNT9
ID   Q9ZNT9      PRELIMINARY;      PRT;   692 AA.
AC   Q9ZNT9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   FERTILIZATION-independent seed 2 protein.
```

```
GN   FIS2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99093530; PubMed=9874812;
RA   Luo M., Bilodeau P., Koltunow A., Dennis E.S., Peacock W.J.,
RA   Chaudhury A.;
RT   "Genes controlling fertilization-independent seed development in
RT   Arabidopsis thaliana.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:296-301(1999).
DR   EMBL; AF096096; AAD09105.1; -.
DR   EMBL; AF096095; AAD09104.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
FT   VARIANT     637    637       W -> *.
SQ   SEQUENCE   692 AA;  77666 MW;  2A3F61B961676497 CRC64;

  Query Match             4.9%;  Score 190;  DB 10;  Length 692;
  Best Local Similarity   18.8%;  Pred. No. 6.8e-05;
  Matches  126;  Conservative   99;  Mismatches  267;  Indels  178;  Gaps   24;

Qy        158 QLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKV----------CHKKRKDVSCPIR 204
              ||:   |  ||       ::  |   | |||:::             |    :|| | |
Db         74 QLSPLTFCSKNRNQRRQRDDSNNVKKLNVLLMELDLDDLPRGTENDSTHVNDDNVSSPPR 133

Qy        205 --------------QVPTGKKQVPLIPDLNQTKPGNF--PSLAVSSNE------------ 236
                           |:   :   | :| :|    ||   |  | || |
Db        134 AHSSEKISDILTTTQLAIAESSEPKVPHVND---GNVSSPPRAHSSAEKNESTHVNDDDD 190

Qy        237 -FEPSNSHMVKSYSLLF----RVTRPGRREFNGMINGETNENIDVNEE---LPARRKRNR 288
               |  :| ::           :: | :       : : ||:  :|:|    | | : ::
Db        191 VSSPPRAHSLEKNESTHVNEDNISSPPKAH-----SSKKNESTHMNDEDVSFPPRTRSSK 245

Qy        289 EDGEKTFVAQMTVFDK--------NRRLQLLDGEYEVAMQEMEECPISKKRATWETILDG 340
              |  :    |  : :         :|| ||    |             | | |   | :
Db        246 ETSDILTTTQPAIVEPSEPKVRRGSRRKQLYAKRY------------KARETQPAIAES 292

Qy        341 KRLPPF----ETFSQGP------------TLQFTLRWTGE-----TNDKSTAPIAKPLA 378
                  | |    |  :  : |          ||:: :   :   :
Db        293 SEPKVLHVNDENVSSPPEAHSLEKASDILTTTQPAIAESSEPKVPHVNDENVSSTPRAHS 352

Qy        379 TRNSESLHQ--ENKPGSVKPTQTIAVKESLTTDLQTRKEKDTP------NENRQKLRIFY 430
              :: ::|  :  :| |   |   :    : |||   |  |   |       ::|       :
Db        353 SKKNKSTRKNVDNVPSPPKTRSSKKTSDILTTTQPTIAESSEPKVRHVNDDNVSSTPRAH 412

Qy        431 QFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVS 490
                | :||:     |::  |  | : :|  ::|     :             |  |  :
Db        413 SSKKNKSTRKND---DNIPSPPKTRSSKKTSNILTRTQ------------PAIAESEPK 456

Qy        491 INECYDGSYAGNP--------QDIHRQPGFA---------------FSRNGPVKRTPITH 527
              :    |   :  |        :  |::   |               :   | |  |
Db        457 VPHVNDDKVSSTPRAHSSKKNKSTHKKDDNASLPPKTRSSKKTSDILATTQPAKAEP--- 513

Qy        528 ILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEM--EVDSED 585
                 || |: |   : | :|  |·::          | ::|| |  |:  :::    |||:
Db        514 ---SEPKVTRVSRRKELHAERCEAKRLERLK---GRQFYHSQTMQPMTFEQVMSNEDSEN 567

Qy        586 EKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVE---- 641
              | |   |       ::|    |:: ||  | |||: | |   |||  :  ||  | :
Db        568 ETDDYALDISERLRLERLVGVSKEEKRYMYLWNIFVRKQRVIADGHVPWACEEFAKLHKE 627

Qy        642 --------NYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGES 693
                     :: .:: :  |  | ::   :  |     | : : |:|         :   | :
Db        628 EMKNSSSFDWWWRMFRIKLWNNGLICAKTFHKCTTILLSNSDEA--------GQFTSGSA 679

Qy        694 ASPANEEITE 703
              |:  |::  |
Db        680 ANANNQQSME 689
```

```
RESULT 11
Q9ZQP0
ID   Q9ZQP0      PRELIMINARY;      PRT;   632 AA.
AC   Q9ZQP0;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Fertilization-independent seed 2 protein.
GN   AT2G35670.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA   Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA   Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA   Fraser C.M., Venter J.C.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006068; AAD15448.2; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ   SEQUENCE   632 AA;  71130 MW;  DDFDE33DC6620AE8 CRC64;

  Query Match             4.6%;  Score 180;  DB 10;  Length 632;
  Best Local Similarity   18.5%;  Pred. No. 0.0003;
  Matches  120;  Conservative   92;  Mismatches  243;  Indels  192;  Gaps   22;

Qy        158 QLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKV----------CHKKRKDVSCPIR 204
              ||:   |  ||       ::  |   | |||:::             |    :|| | |
Db         74 QLSPLTFCSKNRNQRRQRDDSNNVKKLNVLLMELDLDDLPRGTENDSTHVNDDNVSSPPR 133

Qy        205 --------------QVPTGKKQVPLIPDLNQTKPGNF--PSLAVSSNE------------ 236
                           |:   :   | :| :|    ||   |  | || |
Db        134 AHSSEKISDILTTTQLAIAESSEPKVPHVND---GNVSSPPRAHSSAEKNESTHVNDDDD 190

Qy        237 -FEPSNSHMVKSYSLLF----RVTRPGRREFNGMINGETNENIDVNEE---LPARRKRNR 288
               |  :| ::           :: | :       : : ||:  :|:|    | | : ::
Db        191 VSSPPRAHSLEKNESTHVNEDNISSPPKAH-----SSKKNESTHMNDEDVSFPPRTRSSK 245

Qy        289 EDGEKTFVAQMTVFDK--------NRRLQLLDGEYEVAMQEMEECPISKKRATWETILDG 340
              |  :    |  : :         :|| ||    |             | | |   | :
Db        246 ETSDILTTTQPAIVEPSEPKVRRVSRRKQLYAKRY------------KARETQPAIAES 292

Qy        341 KRLPPF----ETFSQGP------------TLQFTLRWTGE-----TNDKSTAPIAKPLA 378
                  | |    |  :  : |          ||:: :   :   :
Db        293 SEPKVLHVNDENVSSPPEAHSLEKASDILTTTQPAIAESSEPKVPHVNDENVSSTPRAHS 352

Qy        379 TRNSESLHQ--ENKPGSVKPTQTIAVKESLTTDLQTRKEKDTP------NENRQKLRIFY 430
              :: ::|  :  :| |   |   :    : |||   |  |   |       ::|       :
Db        353 SKKNKSTRKNVDNVPSPPKTRSSKKTSDILTTTQPTIAESSEPKVRHVNDDNVSSTPRAH 412

Qy        431 QFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVS 490
                | :||:     |::  |  | : :|  ::|
Db        413 SSKKNKSTRKND---DNIPSPPKTRSSKKTSNIL------------------------- 443

Qy        491 INECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESEDGE 550
                          :  | |  |              || |: |  : | :|  |
Db        444 ------------------------ATTQPAKAEP------SEPKVTRVSRRKELHAERCE 473

Qy        551 VEQQRTYSSGHNRLYFHSDTCLPLRPQEM--EVDSEDEKDPEWLREKTITQIEEFSDVNE 608
                :          | ::|| |  |:  :::     |||:| |   |       ::|    |::
Db        474 AKRLERLK---GRQFYHSQTMQPITFEQVMSNEDSENETDDYALDISERLRLERLVGVSK 530

Qy        609 GEKEVMKLWNLHVMKHGFIADNQMNHACMLFVE-----------NYGQKIIKKNLCRNF 656
               ||  | |||: | |   |||  :  ||  | :             ::  :: :  |  |
Db        531 EEKRYMYLWNIFVRKQRVIADGHVPWACEEFAKLHKEEMKNSSSFDWWWRMFRIKLWNNG 590
```

```
Qy        657 MLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITE 703
              ::   : |    | : : |:|         :   | :|:  |::  |
Db        591 LICAKTFHKCTTILLSNSDEA--------GQFTSGSAANANNQQSME 629


RESULT 12
Q8AWB3
ID   Q8AWB3      PRELIMINARY;      PRT;   570 AA.
AC   Q8AWB3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Keratin gamma 2.
OS   Lampetra fluviatilis (River lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Lampetra.
OX   NCBI_TaxID=7748;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schultess J., Schaffeld M., Markl J.;
RT   "Unusual intermediate filament proteins from Lampetra fluviatilis.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ308111; CAC87101.1; -.
SQ   SEQUENCE   570 AA;  58991 MW;  FFAE0FFD3DD2C872 CRC64;

  Query Match             4.4%;  Score 171;  DB 13;  Length 570;
  Best Local Similarity   21.8%;  Pred. No. 0.0011;
  Matches  106;  Conservative   61;  Mismatches  162;  Indels  158;  Gaps   21;

Qy          7 GGGGGGGSGPSAGSGGGGF-------------GGSAAVAAATASGGKSGGGSCGGGGSY 52
              |||| || | | | |||||              ||:| :    |  |  |||  |||| :
Db         47 GGGGCGGGGGSLGGGGGGFALSSYGRGYGAGGGGAAGLGFGLAQFGGGGGGGGGGGGGF 106

Qy         53 SASSS-------------SSAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRF 99
               :|:|             :|||  |   ||
Db        107 GSSASFQVGRSMASGGLHGAAAAFGGGRSLP--------------------------- 137

Qy        100 LRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQL 159
                      |:        ::    || ::    :    : ||::  ::|:  | |||:
Db        138 --------PL--------VTREGQRTVLRSLNDRFAGYIEKVQRL--QEENVQLEAHLK- 178

Qy        160 TFTGFFHKNDKPSPNSE---------------------NEQNSVTLEVLLVKVCHKKRK 197
              : ||    :  |   ||                     | |  : |: |       | |
Db        179 SLTG--GVSVAPVDTSELAGDPTARFVELRAEFDALVLANVQLEIELDNLRAFAQEMKAK 236

Qy        198 DVSCPIRQVPTG-KKQVPLIPDLNQTKPGNF-------PSLAVSSNEFEPSNSHMVKSYS 249
                    :  | | |:       :|  ||         | |   | |  | :::
Db        237 ------YEFEVGVKYQL-------ETDIGNMRKDLDGATELKV---ELESRYSMLLEELE 280

Qy        250 LLFRV---------TRPGRREFNGMINGETNENI-------DVNEELPARRKRNREDGEK 293
               | |           :| : : :| |: : : :|       |:  |  |  ::|| | :
Db        281 FLKRAQAEELLKLGAKPDKVDSSGAIDVDDSRSIDLAGMLGDMRAEYEAIAQKNRADADA 340

Qy        294 TFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGP 353
               |  ||   :::        | | |  |: |   | :    |  | |         |:
Db        341 YFQNQMATIEEDH--HKASSETESATSEITEVQRSMQELQLE--LQG-------LLSKNG 389

Qy        354 TLQFTL-----RWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTT 408
               |  |     |: | |   :  :|    | :|  :  : :   |   | : :| :| |
Db        390 QLHNMLAEVEGRYRFEINTMQSR-VADMDGTLHSSKIELQRQ--LVAFQQLLDIKIALDT 446

Qy        409 DLQTRKE 415
              :: | |:
Db        447 EIATYKQ 453


RESULT 13
Q9MAK1
ID   Q9MAK1      PRELIMINARY;      PRT;  1745 AA.
AC   Q9MAK1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   F27F5.16.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA   Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA   Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA   Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA   Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA   Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA   Ecker J.R.;
RT   "Genomic sequence for Arabidopsis thaliana BAC F27F5 from chromosome
RT   I.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007915; AAF69170.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
SQ   SEQUENCE   1745 AA;  194607 MW;  26BD44B87389E2F9 CRC64;

  Query Match             4.4%;  Score 169.5;  DB 10;  Length 1745;
  Best Local Similarity   20.3%;  Pred. No. 0.0063;
  Matches  124;  Conservative   74;  Mismatches  218;  Indels  195;  Gaps   24;

Qy          8 GGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAAAAGAA 67
              ||  || |    ||| | ||          |||   ||  || |   : :       |
Db        423 GGPSGGDGEGGPSGGDGEGG--------PSGGDGEGGPSGGDGEGGPNGADGEGGPNG-- 472

Qy         68 VLPVKKPKMEHVQADHELFLQAFEKPTQIYR--FLRTRNLI------APIFLHRTL---- 115
                        || |:  :||:|     |  | |   |:        :|:   :
Db        473 -------------ADGEVGDEAFDKYAGELRSFFKRQTQLLDKKFEDLKVFVREEIRAAL 519

Qy        116 ---------TYMSHRNSRTN--IKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGF 164
                      ||::   :  ::| |  |: :  :|||     |  :     :
Db        520 QSQDRKEEGPRSSHKSDSPSEKVERVTKTVEKVNKRVEKANKRVEKAAEQVQRKSV---- 575

Qy        165 FHKNDKPS---PNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVP-------TGKKQVP 214
                |: ||    |  :  |:              ||    | |: :||          ||
Db        576 -KKSTKPRKFVPRRSSRLNNTP----------KKAATGSLPVEEVPGNDGDGEVSAKQTV 624

Qy        215 LIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLF------------RVTRPGRRE 261
                |:     |     | |: | |  || : :  | |                ||  |  :
Db        625 SDYDMENIYLGG----ASSTEEEEGGNSEIEEDSSKLHQNVSDSVMEDNVDRVPSPSHQP 680

Qy        262 FNGMINGETNENI---------------DVN---------------EELPARRKRNRE 289
              :|: :|   | :               ||:               ||: |   :  |
Db        681 EHGIPDGGNFEQLLPDPILSDPALEKMSDDVSSSSHQDVQKGLGGNPEEEVVAPESQQLE 740

Qy        290 DGEKTFVAQMTVF------DKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRL 343
              : :   |     |      | |   : : || | |::|       |     :  |:   |
Db        741 EPQSPEVKVRQEFAFEVTEDGNPSAETICGEGEEALKE------DKSPTVVDEALEDTAL 794

Qy        344 PPFETFSQGPTLQFTLRWTGETNDKS--TAPIAKPLATRNSESLHQENKPGSVKPTQTIA 401
              | | :    ||      :    :| |  |  ::| | |:  : :  ||     |  : :|
Db        795 PGFVS----PTTVGDFNYEANLSDPSSPTVVVSKVL-TQLKDDILAENVS---KIPEKVA 846

Qy        402 VKESLTTDL----------------QTRKEKD------------TPNENRQKLRIFYQ 431
              | | : | |                :| || |            |   ::| :  :
Db        847 VPEEVLTQLKDDVLEEKVSEKVAIPEETGKEDDVVEAGVSKTEAIVAPVNEKEKKSVRRR 906

Qy        432 FLYNNNTRQQ-------------------TEARDDLHCPWCTLN---CRKLYSLLKHLK 468
                :::::|:                  :| | |: |   ::     :|| :: :  |
Db        907 VTFSDDTKNDDKKDPDVTFVEEREAGPEVLVKAGDSLYNPLEKVDEAKFQKLCTITQGEK 966

Qy        469 LCHSRFIFNYV 479
               ||: | |  |
Db        967 RCHTLFGFRKV 977
```

```
RESULT 14
O62007
ID   O62007      PRELIMINARY;      PRT;   620 AA.
AC   O62007;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Intermediate filament protein D1.
OS   Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7740;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98267218; PubMed=9602172;
RA   Riemer D., Karabinos A., Weber K.;
RT   "Analysis of eight cDNAs and six genes for intermediate filament (IF)
RT   proteins in the cephalochordate Branchiostoma reveals differences in
RT   the IF multigene families of lower chordates and the vertebrates.";
RL   Gene 211:361-373(1998).
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR   EMBL; AJ223579; CAA11446.1; -.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR002957; Keratin_I.
DR   InterPro; IPR003054; Keratin_II.
DR   InterPro; IPR000772; Ricin_B_lectin.
DR   Pfam; PF00038; filament; 1.
DR   Pfam; PF00652; Ricin_B_lectin; 3.
DR   PRINTS; PR01248; TYPE1KERATIN.
DR   PRINTS; PR01276; TYPE2KERATIN.
DR   SMART; SM00458; RICIN; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Coiled coil; Intermediate filament.
SQ   SEQUENCE   620 AA;  67101 MW;  83EAB189BF80A91B CRC64;

  Query Match             4.3%;  Score 166.5;  DB 5;  Length 620;
  Best Local Similarity   23.0%;  Pred. No. 0.0026;
  Matches   99;  Conservative   61;  Mismatches  160;  Indels  111;  Gaps   19;

Qy          7 GGGGGGGSGPSAGSGGGG----FGGS----------------AAVAAATASGGKSGGGS 45
              |||||||   |  |||||    ||||                :::  : | ||: |||
Db         31 GGGGGGGYSSSFSSGGGGGRASFGGSSRFGGGGGCGGRTSMSRSSMTRSGAGGGRMGGGG 90

Qy         46 CGGGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHEL------FLQAFEKPTQIYRF 99
               ||||  |   ||  |  |  |:: : : ::|      ||      |    ||     ||
Db         91 -GGGGGRSYGFSSMTAEQASQALVALGQVRVERTGDKDELVGLNDRFATFIEK----VRF 145

Qy        100 LRTRNLIAPIFLHRTLTYMSHRNSRTNI----KRKTFKVDDMLSKVEKMKG--EQESHSL 153
              |  :|      |   |  :  : |  ::    : :  ::  ::  |   :|  | |   |
Db        146 LENQNR----KLEMKLKMVQQKGSGPDLGAMWEAELRQIRQLIEVVNTERGSLEAERDGL 201

Qy        154 SAHLQLTFTGFFHKNDKPSPNSENEQNSVTL-----EVLLVKVCHKKRKDVSCPIRQVPT 208
              |  ::   | :   :|:       |:    :     |  | :|  : | |         :
Db        202 SGEVKELKTRY---DDEVGTRDGLEEEIKKIRADFDEASLTRVDLEARLD---------S 249

Qy        209 GKKQVPLIPDLNQTKPGNFPSLAVSSN--EFEPSNSHMVKSYSLLFRV-------TRPGR 259
              | ::  : ::  :     |  : |   | |||    |  | |  |          || |
Db        250 IKSEIEFLKEVYAAEIEALNSQILDSTMIELEPSAGPDVDLDSCLAEVKAQYEQLTRMSR 309

Qy        260 REFNGMI----------NGETNENI--------------DVNEELPARRKRNR------ 288
              |              :|: | ::               :  |: : : |||
Db        310 AEAESWYATKFEDLQRNSGKNNNDLADARSELSKYQNQIARLQSEIESMKNRNRQLEGQL 369

Qy        289 ---EDGEKTFVA--QMTVFDKNRRLQLLDGEYEVAMQEMEE-----CPISKKRATWETIL 338
                |:  |: :|  |  :      || | ||    |:| :|       :  : | :  :|
Db        370 KNVEESGKSKLADKQAEIEALEAELQRLRGEISKQMREYQELHNVKMALDVEIAAYRKLL 429

Qy        339 DGK--RLPPFE 347
              :|:  ||  |:
Db        430 EGEESRLHGFD 440
```

```
RESULT 15
Q8I519
ID   Q8I519      PRELIMINARY;      PRT;  2359 AA.
AC   Q8I519;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Erythrocyte membrane protein 1 (PfEMP1).
GN   PFL1960W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014850; AAN36478.1; -.
SQ   SEQUENCE   2359 AA;  261601 MW;  29401D4450AD4DB5 CRC64;

  Query Match             4.2%;  Score 165;  DB 5;  Length 2359;
  Best Local Similarity   20.0%;  Pred. No. 0.019;
  Matches   94;  Conservative   65;  Mismatches  172;  Indels  138;  Gaps   18;

Qy          7 GGGGGGGSGPSAGSGG----GGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSAAA 62
              | |||||||   ||||    || |||      :   ||  |||  ||||       |
Db        631 GSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGG 690

Qy         63 AAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRN 122
                :|:             ::| :   : ::                  : |  :  :  :
Db        691 GSGS------------KSDSQKLYEEWK----------------CYKHDEVQKVQGQG 720

Qy        123 SRTNIKRKTFKVDDMLS---------KVEKMKGEQESHSLSAHLQLTFTGFF-------- 165
                : :  : ||: |           :| : | :|  :     | ||  ||
Db        721 EEED-EDEVEEEDDLKSAGGLCILPNPKKKEESEAKSQNEPEQFQKTFYDFFTYWVAHML 779

Qy        166 --------HKNDKPSPNSENEQNSVTLEVLLVKVCHKKRK-DVSCPIRQVPTGKKQVPLI 216
                     | ||   |:  :|           | |  | |  | :: |   | :  |
Db        780 KDSIHWRTEKLDKCINNTNGKQTK----------CRKGCKGDCDCFLKWVNEKKTEWGKI 829

Qy        217 PDLNQTKPGNFPSLAVSSNEFE---PSNSHMVKSYSLLFRVTRPG---RREFNGMINGET 270
              |  :|: |          || |   | ::    :: ::      |     :|    | |
Db        830 KDHFKTQEG-------LDNEGENGIPESTRFGFTHDVVL-----GEFLELQFLNENNEEN 877

Qy        271 NENIDVNEEL-------PARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEME 323
              ::|   :||:          :|:|:: |      : |: |      :|:| |  :|     :
Db        878 SKNSLDSEEIKHLKNLRDIIKKKNQQAGAGA-TGKRTIMD-----ELIDYEEGIAKDCQQ 931

Qy        324 ECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLAT---- 379
              :|  ::|    :  |      | :  |                      || | |: :
Db        932 KCQETQKPQQVDRDLARSATGPRDPSS----------------PAAPTAGPVESAAVD 973

Qy        380 -------RNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNE 421
                    |   :|::   |  :  :  ||        ::|:  | |
Db        974 DEDDDNEEEEEEEEEDEKEEVGGSDVVEEKEE-------KEEEKEPEE 1015


Search completed: September  2, 2003, 11:04:15
Job time : 60.1833 secs
```

```
              ||:   | |:       |  : | |     | |        :::       |: :  |
Db        238 NDGNRGLGYPEATELAGQFEMTSNIP---PAIAHSSLDAGAKVILTSEAVVPATKTRKLS 294

Qy        546 SEDGEVEQ----QRTYSSGH----NRLYFHSDTCLPLRPQEM--EVDSEDEKDPEWLREK 595
              :|  |           || |     | ::||    |:  :::  : |||||  | :    :
Db        295 AERSEARSTAILTEKISSSHLLLQKRQFYHSHRVQPMALEQVMSDRDSEDEVDDDVADFE 354

Qy        596 TITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIK-KNLCR 654
                 :::| |||: ||: | |||  | |   |||  :: ||  |   | ::: :  :|
Db        355 DRQMLDDFVDVNKDEKQFMHLWNSFVRKQRVIADGHISWACEAFSRFYEKELHRYSSLFW 414

Qy        655 NFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGF-- 712
              : | |: : :  |:   :|:   | |   :   :   : :  : :   : |   |
Db        415 CWRLFLIKLWNHGLVDSATINNCNTILENWRNSSDTTTTNNNNSVDRPSDSNTNNNNIVS 474

Qy        713 --SEINSKEKALETDSVS 728
              ::||:|      |: |
Db        475 HPNDINNKNNVDNKDNNS 492


Search completed: September  2, 2003, 11:06:18
Job time : 93.058 secs
```

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 230; DB 21; Length 498;
Best Local Similarity 22.5%; Pred. No. 1.3e-09;
Matches 85; Conservative 62; Mismatches 161; Indels 70; Gaps 12;

```
Qy      413 RKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHS 472
            || |     :|    : : :   ||| |:|| |:|  ||:|:: |     |  ||   |
Db      123 RKRK-----SRSTGMVVFNYKDCNNTLQKTEVREDCSCPFCSMLCGSFKGLQFHLNSSHD 177

Qy      473 RFIFNYVY----------------HPKGARIDV-SINECY------------------ 495
              | | : |                    :| | :   :  |:
Db      178 LFEFEFKYISVNPTIMFCSSKPRKRRQRGGRNNTRRLKVCFLPLDSPSLTNGTENGITLL 237

Qy      496 -DGSYA-GNPQDIHRQPGFAFSRNGPVKRTPITH--------ILVCRPKRTKASMSEFLE 545
```

```
PT   Novel DNA used to produce transgenic plants with altered floral
PT   regulation which can have increased crop yields -
XX
PS   Claim 2; Page 12-15; 53pp; English.
XX
CC   The present sequence represents a protein having a flowering regulating
CC   activity, which is designated MPC1. The genomic sequence is given
CC   in AAZ36948. A rice MPC1 is also disclosed in the specification. The
CC   rice and Arabidopsis cDNAs show significant homology with each other.
CC   A naturally occurring mutation of the MPC1 gene eliminates normal
CC   flowering regulating ability of plants, and leads to flowering
CC   immediately after germination (super early flowering mutation), and
CC   leads to a truncated protein comprising amino acids 1-540 of the present
CC   sequence. The MPC1 polynucleotide sequence can be used to produce plants
CC   with altered flowering times in comparison with wild type plants, by
CC   enhancing or inhibiting the expression of the flowering regulating gene.
CC   Antisense polynucleotides can be used to reproduce the effects of the
CC   mutated MPC1 gene. This alteration can be used to increase the yield
CC   of food crops. Flowering inhibition of vegetables increases their
CC   productivity.
XX
SQ   Sequence   611 AA;

  Query Match             6.1%;  Score 236;  DB 21;  Length 611;
  Best Local Similarity   19.4%;  Pred. No. 5.8e-10;
  Matches  137;  Conservative  104;  Mismatches  258;  Indels  208;  Gaps   24;

Qy         77 EHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDD 136
              | : |:    | |: || ::|  :: | :  |:|| | | |      :  |:   |
Db         14 EEIAAEES--LAAYCKPVELYNIIQRRAIRNPLFLQRCLHYKIEAKHKRRIQMTVFL--- 68

Qy        137 MLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKR 196
                        | :: | :|                    |    | :|| ::   |
Db         69 -------------SGAIDAGVQ------------------TQKLFPLYILLARLVSPK--95

Qy        197 KDVSCPIRQ-----------VPTGKKQVP---------LIPDLN----QTKPGN----FP 228
                 |: :           : ||   |          |:||:|    : | |:    |
Db         96 -----PVAEYSAVYRFSRACILTGGLGVDGVSQAQANFLLPDMNRLALEAKSGSLAILFI 150

Qy        229 SLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNR 288
              | | : |     :|  : |               | | |       :  ::| :
Db        151 SFAGAQNSQFGIDSGKIHS---------------GNIGGHC-----LWSKIPLQ----- 184

Qy        289 EDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFET 348
                    :    : |:  : |      |:: ||: | |  |  : |  :
Db        185 --------SLYASWQKSPNMDLGQRVDTVSLVEMQPCFIKLKSMSEEKCVS--------- 227

Qy        349 FSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTT 408
               | |:   |          |::|    :     |    |  | |      |:   ||
Db        228 -IQVPSNPLT----------SSSPQQVQVTISAEEVGSTEKSPYSSFSYNDIS-SSSLLQ 275

Qy        409 DLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLK 468
              :: |    | |       : : : | ||  |:||  :|  ||:| : |     |  ||
Db        276 IIRLR----TGN-------VVFNYRYYNNKLQKTEVTEDFSCPFCLVKCASFKGLRYHLP 324

Qy        469 LCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIH-RQPGFAF------------- 514
               |    | :    :   ::||:      |   |  |:  :|  | |
Db        325 STHDLLNFEFWVTEEFQAVNVSLKTETMISKV-NEDDVDPKQQTFFFSSKKFRRRRQKSQ 383

Qy        515 ---SRNGP--------VKRTPITHILVCRPKR--------------------------- 535
                 || ||         : :|   | :      |
Db        384 VRSSRQGPHLGLGCEVLDKTDDAHSVRSEKSRIPPGKHYERIGGAESGQRVPPGTSPADV 443

Qy        536 ------------TKASMSEFLESEDGEVEQQ--RTYSSGHNRLYFHSDTCLPLRPQEM-- 579
                         ::| :| ::    :|:   |  |    | :|||    |:  :::
Db        444 QSCGDPDYVQSIAGSTMLQFAKTRKISIERSDLRNRSLLQKRQFFHSHRAQPMALEQVLS 503

Qy        580 EVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLF 639
              : ||||| | :    :    :::| || : ||::| :||  | |  :||  :  ||  |
Db        504 DRDSEDEVDDDVADFEDRRMLDDFVDVTKDEKQMMHMWNSFVRKQRVLADGHIPWACEAF 563

Qy        640 VENYGQKIIK-KNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQ 685
                 :|  :::  :|   : : :| : :  |:   :::   | | ::|
```

```
Db        564 SRLHGPIMVRTPHLIWCWRVFMVKLWNHGLLDARTMNNCNTFLEQLQ 610


RESULT 15
AAG54168
ID   AAG54168 standard; Protein; 498 AA.
XX
AC   AAG54168;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 69039.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
```

```
ID   AAY53933 standard; Protein; 604 AA.
XX
AC   AAY53933;
XX
DT   13-MAR-2000  (first entry)
XX
DE   An Os-MPC1 protein having flowering regulating activity.
XX
KW   Flowering regulating activity; MPC1; flowering; germination;
KW   super early flowering mutation; altered flowering time;
KW   flowering regulating gene; food crop; vegetable; flowering inhibition;
KW   productivity.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   Domain          310..335
FT                   /note= "zinc finger domain"
XX
PN   EP967278-A2.
XX
PD   29-DEC-1999.
XX
PF   28-JUN-1999;   99EP-0305077.
XX
PR   26-JUN-1998;   98JP-0180065.
PR   24-JUN-1999;   99JP-0179043.
XX
PA   (MITA ) MITSUI CHEM INC.
XX
PI   Yoshida N,  Kato Y,  Takahashi S,  Yanai Y,  Hiratsuka J,  Miwa T;
XX
DR   WPI; 2000-064612/06.
DR   N-PSDB; AAZ36953.
XX
PT   Novel DNA used to produce transgenic plants with altered floral
PT   regulation which can have increased crop yields -
XX
PS   Claim 3; Page 36-39; 53pp; English.
XX
CC   The present sequence represents a protein having a flowering regulating
CC   activity, which is designated Os-MPC1. An Arabidopsis MPC1 is also
CC   disclosed in the specification. The rice and Arabidopsis cDNAs show
CC   significant homology with each other. A naturally occurring mutation
CC   of the MPC1 gene eliminates normal flowering regulating ability of
CC   plants, and leads to flowering immediately after germination (super
CC   early flowering mutation). The MPC1 polynucleotide sequence can be
CC   used to produce plants with altered flowering times in comparison
CC   with wild type plants, by enhancing or inhibiting the expression of
CC   the flowering regulating gene. Antisense polynucleotides can be used
CC   to reproduce the effects of the mutated MPC1 gene. This alteration
CC   can be used to increase the yield of food crops. Flowering inhibition
CC   of vegetables increases their productivity.
XX
SQ   Sequence   604 AA;

  Query Match             6.3%;  Score 243.5;  DB 21;  Length 604;
  Best Local Similarity   19.1%;  Pred. No. 1.4e-10;
  Matches  134;  Conservative   97;  Mismatches  249;  Indels  223;  Gaps   23;

Qy         77 EHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH--RNSRTNI------- 127
              | : |:    |  : || ::|  :: |::  | || | | |  |  |  |: |
Db         14 EQLAAEESFAL--YCKPVELYNIIQRRSIKNPAFLQRCLLYKIHARRKKRSLITISLSGG 71

Qy        128 KRKTFKVDDMLSKVEKMKGEQESHSLSAH----------LQLTFTGFFHKNDKPS----P 173
                |  :  ::      :     : ||  |          |  :|  | :|:   :    |
Db         72 TNKELRAQNIFPLYVLLARPTNNVSLEGHSPIYRFSRACLLTSFHEFGNKDYTEATFVIP 131

Qy        174 NSEN--EQNSVTLEVLLVKVCHKKRK---DVSCPIRQVPTGKKQ----------VPLIPD 218
              : :|     : :| ::|:  | :  :   | :|    |      |          :|:  |
Db        132 DVKNLATSRACSLNIILIS-CGRAEQTFDDNNCSGNHVEGSTLQKLEGKCFWGKIPI--D 188

Qy        219 LNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNE 278
                |  :  ||  ||::                                  | | |     |
Db        189 LLASSLGNCVSLSL---------------------------------GHTVEMSSTVE 213

Qy        279 ELPA--RRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWET 336
                |:    |   :|   || :|      : :||:     |   ::| | | |
Db        214 MTPSFLEPKFLEDDSCLTFCSQKVDATGSFQLQVSISAQEAGAKDMSESPYS-------- 265

Qy        337 ILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKP 396
              :     :||
Db        266 VYSYNDVPP---------------------------------------------------- 274

Qy        397 TQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLN 456
                    |||  :: |              : : : | ||| |:|| :|  :|  ||:| :
Db        275 -------SSLTHIIRLRSG----------NVLFNYKYYNNTMQKTEVTEDFSCPFCLVP 316

Qy        457 CRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSI-NECYDGSYAGNPQDIHRQPGFAFS 515
              |     |  ||   |  | : :    :   ::||:  : :          |   |     |
Db        317 CGSFKGLGCHLNASHDLFHYEFWISEECQAVNVSLKTDSWRTELLAEGVDPRHQTFSYRS 376

Qy        516 RNGPVKRTPIT---------HILVC-RPKRTKASMSEFLESEDGEVEQQRTYSSGH---- 561
              |    ||  |:         ||:    |:  :|       |||  |:::   |  |
Db        377 RFKKRKRVEISSDKIRHVHPHIVDSGSPEDAQAG------SEDDYVQRENGSSVAHASVD 430

Qy        562 ------------------------------------NRLYFHSDTCLPLRPQEM--EV 581
                                                   | :|||     |:  ::  :
Db        431 PANSLHGSNLSAPTVLQFGKTRKLSVERADPRNRQLLQKRQFFHSHRAQPMAWSKVFSDR 490

Qy        582 DSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVE 641
              ||||| | :    :    :::| || : || :| :|| | |   :|| :  ||  | :
Db        491 DSEDEVDDDIADFEDRRMLDDFVDVTKDEKLIMHMWNSFVRKQRVLADGHIPWACEAFSQ 550

Qy        642 NYGQKIIKKNL---CRNFMLHLVSMHDFNLISIMSIDKAVTKL 681
                :||::::      |  |   :| : : :|:   :::   | |
Db        551 FHGQELVQNPALLWCWRFF--MVKLWNHSLLDARAMNACNTIL 591


RESULT 14
AAY53932
ID   AAY53932 standard; Protein; 611 AA.
XX
AC   AAY53932;
XX
DT   13-MAR-2000  (first entry)
XX
DE   A MPC1 protein having flowering regulating activity.
XX
KW   Flowering regulating activity; MPC1; flowering; germination;
KW   super early flowering mutation; altered flowering time;
KW   flowering regulating gene; food crop; vegetable; flowering inhibition;
KW   productivity.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   Domain          306..327
FT                   /note= "zinc finger domain"
XX
PN   EP967278-A2.
XX
PD   29-DEC-1999.
XX
PF   28-JUN-1999;   99EP-0305077.
XX
PR   26-JUN-1998;   98JP-0180065.
PR   24-JUN-1999;   99JP-0179043.
XX
PA   (MITA ) MITSUI CHEM INC.
XX
PI   Yoshida N,  Kato Y,  Takahashi S,  Yanai Y,  Hiratsuka J,  Miwa T;
XX
DR   WPI; 2000-064612/06.
DR   N-PSDB; AAZ36947, AAZ36948.
XX
```

```
  Matches  165;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        429 FYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARID 488
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FYQXLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARID 60

Qy        489 VSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESED 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESED 120

Qy        549 GEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEF 603
              |||||||||||||||| |||||||||||||||||||||||||| | : ::: |
Db        121 GEVEQQRTYSSGHNRXYFHSDTCLPLRPQEMEVDSEDEKDPEWLRXKPLHKLKSF 175


RESULT 12
ABU55471
ID   ABU55471 standard; Protein; 175 AA.
XX
AC   ABU55471;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #558.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-0764864.
XX
PR   31-JAN-2000; 2000US-179065P.
PR   04-FEB-2000; 2000US-180628P.
PR   28-JUN-2000; 2000US-214886P.
PR   07-JUL-2000; 2000US-216647P.
PR   07-JUL-2000; 2000US-216880P.
PR   11-JUL-2000; 2000US-217487P.
PR   11-JUL-2000; 2000US-217496P.
PR   14-JUL-2000; 2000US-218290P.
PR   26-JUL-2000; 2000US-220963P.
PR   26-JUL-2000; 2000US-220964P.
PR   14-AUG-2000; 2000US-224518P.
PR   14-AUG-2000; 2000US-224519P.
PR   14-AUG-2000; 2000US-225267P.
PR   14-AUG-2000; 2000US-225268P.
PR   14-AUG-2000; 2000US-225270P.
PR   14-AUG-2000; 2000US-225447P.
PR   14-AUG-2000; 2000US-225757P.
PR   14-AUG-2000; 2000US-225758P.
PR   22-AUG-2000; 2000US-226868P.
PR   30-AUG-2000; 2000US-228924P.
PR   01-SEP-2000; 2000US-229287P.
PR   01-SEP-2000; 2000US-229343P.
PR   01-SEP-2000; 2000US-229344P.
PR   01-SEP-2000; 2000US-229345P.
PR   05-SEP-2000; 2000US-229509P.
PR   05-SEP-2000; 2000US-229513P.
PR   08-SEP-2000; 2000US-231413P.
PR   21-SEP-2000; 2000US-234223P.
PR   21-SEP-2000; 2000US-234274P.
PR   25-SEP-2000; 2000US-234997P.
PR   27-SEP-2000; 2000US-235834P.
PR   29-SEP-2000; 2000US-236327P.
PR   29-SEP-2000; 2000US-236367P.
PR   29-SEP-2000; 2000US-236368P.
PR   29-SEP-2000; 2000US-236369P.
PR   29-SEP-2000; 2000US-236370P.
PR   02-OCT-2000; 2000US-236802P.
PR   02-OCT-2000; 2000US-237037P.
PR   02-OCT-2000; 2000US-237038P.
PR   02-OCT-2000; 2000US-237039P.
PR   02-OCT-2000; 2000US-237040P.
PR   13-OCT-2000; 2000US-239935P.
PR   20-OCT-2000; 2000US-240960P.
PR   20-OCT-2000; 2000US-241785P.
PR   20-OCT-2000; 2000US-241809P.
PR   01-NOV-2000; 2000US-244617P.
PR   17-NOV-2000; 2000US-249299P.
PR   08-DEC-2000; 2000US-251856P.
PR   08-DEC-2000; 2000US-251868P.
PR   08-DEC-2000; 2000US-251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73730.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for
PT   treating, inhibiting or preventing e.g. neural, immune system,
PT   muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT   cardiovascular or renal disorders -
XX
PS   Claim 11; SEQ ID NO 1355; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC   arthritis and multiple sclerosis), muscular disorders, respiratory
CC   diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC   reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC   cardiovascular disorders (e.g. congenital heart defects, Ebstein's
CC   anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC   kidney failure and end-stage renal disease), hyperproliferative disorders
CC   (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC   septic shock, bursitis and appendicitis), allergic reactions and
CC   conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC   atherosclerosis and myocardial infarction) and cancerous diseases.
CC   Sequences ABU54914-ABU55699 and ABU55748 represent human novel
CC   polypeptides of the invention.
XX
SQ   Sequence   175 AA;

  Query Match             23.7%;  Score 919;  DB 24;  Length 175;
  Best Local Similarity   94.3%;  Pred. No. 1.3e-65;
  Matches  165;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        429 FYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARID 488
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FYQXLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARID 60

Qy        489 VSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESED 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESED 120

Qy        549 GEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQIEEF 603
              |||||||||||||||| |||||||||||||||||||||||||| | : ::: |
Db        121 GEVEQQRTYSSGHNRXYFHSDTCLPLRPQEMEVDSEDEKDPEWLRXKPLHKLKSF 175


RESULT 13
AAY53933
```

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26389.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 1355; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 23.7%; Score 919; DB 22; Length 175;
Best Local Similarity 94.3%; Pred. No. 1.3e-65;

```
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15738.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 41697; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   955 AA;

  Query Match             30.3%;  Score 1177;  DB 22;  Length 955;
  Best Local Similarity   34.8%;  Pred. No. 3e-85;
  Matches  281;  Conservative  136;  Mismatches  225;  Indels  166;  Gaps   26;

Qy         43 GGSCGG--GGSYSASSSSSAAAAAGAAVLPVK--KPKMEHVQADHELFLQAFEKPTQIYR 98
              |:  |  | :: |  :|:    ||: | |    : |:   | : ||||||||||||||
Db         15 GSAANGIIGLTHGAPDASN----AGSTVPPTAEGQVKLNGHQQEQELFLQAFEKPTQIYR 70

Qy         99 FLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSA-HL 157
              :|| |:   ||||:|||:||  | || | || :|:|: ||  :      |:| ::|  :|
Db         71 YLRNRHETNPIFLNRTLSYMKERMSRNNKKRISFQVNSMLESI-----TQKSEAVSQNYL 125

Qy        158 QLTFTGFFHKNDKPSPNSENEQ----------NSVTLEVLLVKVCHKKRKDVSCPIRQV 206
               : :     |      |   |             ||::|  | |:   |||| :   ::: 
Db        126 HVIYDSLHEKLPARLDNESGEDLLQEQLLCEAGESVSVETTLYKITRSKRKDSTLDFQEL 185

Qy        207 PTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEP-SNSHMVKSYSLLFRVTRPGRREFNGM 265
              :   |:   |   : : |   ::::      |   | :  | ||||:          :
Db        186 LSKCSQIVYNP---KDRVGEHATISIPLQTMRPMGEQHTL--YKLLFRIK---------V 231

Qy        266 INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEEC 325
              ::  |    | | | |   ||:| : || | ::: :::|:    : :||||  :| :
Db        232 LSPSTCN--DENAETPP-NKRSRPN-EKMFGSELILYEKSSGF-ITEGEYEAMLQPLNST 286

Qy        326 PI---SKKRATWETILDGKRLP---PFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLAT 379
               |   | |: ||||: |   :|    :: : | | |:| |     :|::    |: |
Db        287 SIKSFSPKKCTWETMPD-SYIPLSLTYDVYQQSPMLKFHLTL---SNEQLPEMISAPELQ 342

Qy        380 R------------NSESLHQEN-----KPGSVKPTQTIAVKESLTTDLQTRKEKDTPNE 421
              |            |: : :  |     | ||     |:                | ||
Db        343 RYVQHLDAVAEMNYNNNNYNNNNNCSGLKNGSGGGNSTVC-------------KTTP-- 386

Qy        422 NRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVYH 481
               : ::| | |:|:||||||||   :|:|||| |:| :||:|||||||||:|| | |
Db        387 --EHIQIVYNFMYSNNTRQQTEYTQELNCPWCGLDCLRLYALLKHLKLCHARFNFTYQPA 444

Qy        482 PKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRN-GPVKRTPITHILVCRPKRTKASM 540
              ||||||:||: |||||||:| |:    | :|:|  |||:|| :| ::||||:| |  :
Db        445 GSGARIDVTINDAYDGSYAGSPYDLAGPSGSSFARTCGPVRRTSVTSLMVCRPRRQKTCL 504

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              |||| :: |:  ||:| :|||||| |::|||: |:|:::||| | || |||:|||  |
Db        505 DEFLELDEDEISNQRSYITGHNRLYHHTETCLPVHPKELDIDSEGESDPLWLRQKTIQMI 564

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              :|||||||||||:|||||||||:|||: | |:  || :|::  | :|::||| |||:||:
Db        565 DEFSDVNEGEKELMKLWNLHVMRHGFVGDCQLPIACEMFLDAKGTEIVRKNLYRNFILHM 624

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGES---------------------------- 693
                |: |: ||:  :: | | ||: :  |   |:                 .
Db        625 CSLFDYGLIAAETVYKTVQKLQGLLSKYAAGQELMQRQREEQLKYWLDVGMHKKQEDPKT 684

Qy        694 -------ASPANEEITE---------------------------------EQNGTANG 711
                    | ||::  |                                    :  || ||
Db        685 LKSPQKPAPPADQASTSSASTSGSGSGSSSMQPPKRMPAHLKRGSAASSPGVQSKGTENG 744

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
               :  ||      : :   |:|:|  | |
Db        745 TNGSNS-----SSSNSKNVAKKSADQPL 767


RESULT 11
AAU16402
ID   AAU16402 standard; Protein; 175 AA.
XX
AC   AAU16402;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1355.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
```

KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-0764864.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR N-PSDB; ABX73744.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for
PT treating, inhibiting or preventing e.g. neural, immune system,
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT cardiovascular or renal disorders -
XX
PS Claim 11; SEQ ID NO 1369; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC arthritis and multiple sclerosis), muscular disorders, respiratory
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC cardiovascular disorders (e.g. congenital heart defects, Ebstein's
CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC kidney failure and end-stage renal disease), hyperproliferative disorders
CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC septic shock, bursitis and appendicitis), allergic reactions and
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC atherosclerosis and myocardial infarction) and cancerous diseases.
CC Sequences ABU54914-ABU55699 and ABU55748 represent human novel
CC polypeptides of the invention.
XX
SQ Sequence 292 AA;

Query Match 37.4%; Score 1453; DB 24; Length 292;
Best Local Similarity 96.8%; Pred. No. 3.5e-108;
Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```
Qy        196 RKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVT 255
              :||||||| |||||| | || ||||||||| |||||||||||||| |||||||||||||
Db          1 KKDVSCPIXQVPTGKXQXPLNPDLNQTKPGXFPSLAVSSNEFEPSXSHMVKSYSLLFRVT 60

Qy        256 RPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEY 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEY 120

Qy        316 EVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAK 375
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 XVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAK 180

Qy        376 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 240

Qy        436 NNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYV 479
              |||||||||||||||||||||||||||||||||||||||||||| |
Db        241 NNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNXV 284
```

RESULT 10
ABB71635
ID ABB71635 standard; Protein; 955 AA.
XX
AC ABB71635;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 41697.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.

PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26403.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 1369; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 37.4%; Score 1453; DB 22; Length 292;
Best Local Similarity 96.8%; Pred. No. 3.5e-108;
Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```
Qy        196 RKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVT 255
              :|||||| |||||| | || ||||||||| |||||||||||||| |||||||||||||
Db          1 KKDVSCPIXQVPTGKXQXPLNPDLNQTKPGXFPSLAVSSNEFEPSXSHMVKSYSLLFRVT 60

Qy        256 RPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEY 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEY 120

Qy        316 EVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAK 375
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 XVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAK 180

Qy        376 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 435
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 240

Qy        436 NNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYV 479
              |||||||||||||||||||||||||||||||||||||||||| |
Db        241 NNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNXV 284
```

RESULT 9
ABU55485
ID ABU55485 standard; Protein; 292 AA.
XX
AC ABU55485;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #572.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;

```
PT   cardiovascular or renal disorders -
XX
PS   Claim 11; SEQ ID NO 911; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC   arthritis and multiple sclerosis), muscular disorders, respiratory
CC   diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC   reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC   cardiovascular disorders (e.g. congenital heart defects, Ebstein's
CC   anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC   kidney failure and end-stage renal disease), hyperproliferative disorders
CC   (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC   septic shock, bursitis and appendicitis), allergic reactions and
CC   conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC   atherosclerosis and myocardial infarction) and cancerous diseases.
CC   Sequences ABU54914-ABU55699 and ABU55748 represent human novel
CC   polypeptides of the invention.
XX
SQ   Sequence   289 AA;

  Query Match            38.5%;  Score 1497;  DB 24;  Length 289;
  Best Local Similarity  99.6%;  Pred. No. 1e-111;
  Matches  283;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        456 NCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFS 515
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 DCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFS 65

Qy        516 RNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLR 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 RNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLR 125

Qy        576 PQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHA 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 PQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHA 185

Qy        636 CMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESAS 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESAS 245

Qy        696 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||||||||||||||||||||||||||
Db        246 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 289


RESULT 8
AAU16416
ID   AAU16416 standard; Protein; 292 AA.
XX
AC   AAU16416;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1369.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
```

```
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 11; SEQ ID No 911; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC   are also used in diagnosing a pathological condition or susceptibility
CC   to a pathological condition. Antibodies to the proteins can also
CC   be used in alleviating symptoms associated with the disorders and in
CC   diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC   include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC   and ocular disorders e.g. corneal infection, and many other
CC   disorders listed in the specification. The polypeptides can also
CC   be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention.

  Query Match             38.5%;  Score 1497;  DB 22;  Length 289;
  Best Local Similarity   99.6%;  Pred. No. 1e-111;
  Matches  283;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        456 NCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFS 515
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 DCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFS 65

Qy        516 RNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLR 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 RNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLR 125

Qy        576 PQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHA 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 PQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHA 185

Qy        636 CMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESAS 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESAS 245

Qy        696 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||||||||||||||||||||||||||||
Db        246 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 289


RESULT 7
ABU55027
ID   ABU55027 standard; Protein; 289 AA.
XX
AC   ABU55027;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #114.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-0764864.
XX
PR   31-JAN-2000; 2000US-179065P.
PR   04-FEB-2000; 2000US-180628P.
PR   28-JUN-2000; 2000US-214886P.
PR   07-JUL-2000; 2000US-216647P.
PR   07-JUL-2000; 2000US-216880P.
PR   11-JUL-2000; 2000US-217487P.
PR   11-JUL-2000; 2000US-217496P.
PR   14-JUL-2000; 2000US-218290P.
PR   26-JUL-2000; 2000US-220963P.
PR   26-JUL-2000; 2000US-220964P.
PR   14-AUG-2000; 2000US-224518P.
PR   14-AUG-2000; 2000US-224519P.
PR   14-AUG-2000; 2000US-225267P.
PR   14-AUG-2000; 2000US-225268P.
PR   14-AUG-2000; 2000US-225270P.
PR   14-AUG-2000; 2000US-225447P.
PR   14-AUG-2000; 2000US-225757P.
PR   14-AUG-2000; 2000US-225758P.
PR   22-AUG-2000; 2000US-226868P.
PR   30-AUG-2000; 2000US-228924P.
PR   01-SEP-2000; 2000US-229287P.
PR   01-SEP-2000; 2000US-229343P.
PR   01-SEP-2000; 2000US-229344P.
PR   01-SEP-2000; 2000US-229345P.
PR   05-SEP-2000; 2000US-229509P.
PR   05-SEP-2000; 2000US-229513P.
PR   08-SEP-2000; 2000US-231413P.
PR   21-SEP-2000; 2000US-234223P.
PR   21-SEP-2000; 2000US-234274P.
PR   25-SEP-2000; 2000US-234997P.
PR   27-SEP-2000; 2000US-235834P.
PR   29-SEP-2000; 2000US-236327P.
PR   29-SEP-2000; 2000US-236367P.
PR   29-SEP-2000; 2000US-236368P.
PR   29-SEP-2000; 2000US-236369P.
PR   29-SEP-2000; 2000US-236370P.
PR   02-OCT-2000; 2000US-236802P.
PR   02-OCT-2000; 2000US-237037P.
PR   02-OCT-2000; 2000US-237038P.
PR   02-OCT-2000; 2000US-237039P.
PR   02-OCT-2000; 2000US-237040P.
PR   13-OCT-2000; 2000US-239935P.
PR   20-OCT-2000; 2000US-240960P.
PR   20-OCT-2000; 2000US-241785P.
PR   20-OCT-2000; 2000US-241809P.
PR   01-NOV-2000; 2000US-244617P.
PR   17-NOV-2000; 2000US-249299P.
PR   08-DEC-2000; 2000US-251856P.
PR   08-DEC-2000; 2000US-251868P.
PR   08-DEC-2000; 2000US-251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73286.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for
PT   treating, inhibiting or preventing e.g. neural, immune system,
PT   muscular, respiratory, reproductive, gastrointestinal, pulmonary,
```

XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS25945.

PF 17-JAN-2001; 2001US-0764864.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR N-PSDB; ABX73306.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for
PT treating, inhibiting or preventing e.g. neural, immune system,
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT cardiovascular or renal disorders -
XX
PS Claim 11; SEQ ID NO 931; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC arthritis and multiple sclerosis), muscular disorders, respiratory
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC cardiovascular disorders (e.g. congenital heart defects, Ebstein's
CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC kidney failure and end-stage renal disease), hyperproliferative disorders
CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC septic shock, bursitis and appendicitis), allergic reactions and
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC atherosclerosis and myocardial infarction) and cancerous diseases.
CC Sequences ABU54914-ABU55699 and ABU55748 represent human novel
CC polypeptides of the invention.
XX
SQ Sequence 388 AA;

Query Match 52.6%; Score 2043; DB 24; Length 388;
Best Local Similarity 99.5%; Pred. No. 2.2e-155;
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        352 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 60

Qy        412 TRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCH 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCH 120

Qy        472 SRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVC 531
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 SRFIFNYVYHPKGARIDVSINEGYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVC 180

Qy        532 RPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEW 591
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        181 RPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHRDTCLPLRPQEMEVDSEDEKDPEW 240

Qy        592 LREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKN 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKN 300

Qy        652 LCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANG 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANG 360

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||||||||||||
Db        361 FSEINSKEKALETDSVSGVSKQSKKQKL 388
```

RESULT 6
AAU15958
ID AAU15958 standard; Protein; 289 AA.
XX
AC AAU15958;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 911.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.

PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS25965.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 931; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 52.6%; Score 2043; DB 22; Length 388;
Best Local Similarity 99.5%; Pred. No. 2.2e-155;
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        352 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 60

Qy        412 TRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCH 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCH 120

Qy        472 SRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVC 531
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 SRFIFNYVYHPKGARIDVSINEGYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVC 180

Qy        532 RPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEW 591
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        181 RPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHRDTCLPLRPQEMEVDSEDEKDPEW 240

Qy        592 LREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKN 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKN 300

Qy        652 LCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANG 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANG 360

Qy        712 FSEINSKEKALETDSVSGVSKQSKKQKL 739
              ||||||||||||||||||||||||||||
Db        361 FSEINSKEKALETDSVSGVSKQSKKQKL 388
```

RESULT 5
ABU55047
ID ABU55047 standard; Protein; 388 AA.
XX
AC ABU55047;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #134.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        309 ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMXEMEECPISKKR 368

Qy        332 ATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 391
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 ATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 428

Qy        392 GSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDLHCP 451
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 GSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYNNNTRQQTEARDDXXXX 488

Qy        452 WCTLNCRKLYSLLKHLKLCHSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPG 511
                                  ||||||||||||||||||||||||||||||||||||||||
Db        489 XXXXXXXXXXXXXXXXXXXXSRFIFNYVYHPKGARIDVSINECYDGSYAGNPQDIHRQPG 548

Qy        512 FAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTC 571
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 FAFSRNGPVKRTPITHILVCRPKRTKASMSEFLESEDGEVEQQRTYSSGHNRLYFHSDTC 608

Qy        572 LPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQ 631
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 LPLRPQEMEVDSEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWNLHVMKHGFIADNQ 668

Qy        632 MNHACMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKG 691
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 MNHACMLFVENYGQKIIKKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKG 728

Qy        692 ESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 739
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db        729 ESASPANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQKL 776


RESULT 4
AAU15978
ID   AAU15978 standard; Protein; 388 AA.
XX
AC   AAU15978;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 931.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
```

```
Qy          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60

Qy         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120

Qy        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180

Qy        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240

Qy        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300

Qy        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360

Qy        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420

Qy        421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480
              |||||||||||||||||||||||||                              |||||||||
Db        421 ENRQKLRIFYQFLYNNNTRQQTEARDDXXXXXXXXXXXXXXXXXXXXXXXXSRFIFNYVY 480

Qy        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540

Qy        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600

Qy        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660

Qy        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720

Qy        721 ALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||
Db        721 ALETDSVSGVSKQSKKQKL 739
```

RESULT 3
ABG30843
ID ABG30843 standard; Protein; 776 AA.
XX
AC ABG30843;
XX
DT 21-OCT-2002 (first entry)
XX
DE Human jAZF1/jjAZ1 fusion protein.
XX
KW Human; jAZF1; juxtaposed with another zinc finger; jjAZ1; jAZF1/jjAZ1;
KW joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;
KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
KW fusion protein
XX
OS Chimeric - Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 13..37
FT /note= "Encoded by ACCTGCCGATTCGGGGGCTGCGGACTCCACTTC-
FT CCCACCCTGGCCGACCTCATCGAGCACATCGAGGACAACCAC.
FT C2H2 zinc finger domain"
FT Region 99..100
FT /note= "Breakpoint for production of fusion protein
FT jAZF1/jJAZ1"
FT Misc-difference 357
FT /note= "Encoded by CAG"
FT Misc-difference 485..508
FT /note= "Encoded by CTGCATTGCCCTTGGTGTACTCTGAACTGCCGCAA-
FT ACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT.
FT C2H2 zinc finger domain"
FT Region 558..575
FT /note= "Bipartite nuclear localisation signal"
XX
PN WO200193805-A2.
XX
PD 13-DEC-2001.
XX
PF 04-JUN-2001; 2001WO-US17936.
XX
PR 02-JUN-2000; 2000US-209093P.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX
PI Koontz J, Sklar J;
XX
DR WPI; 2002-575047/61.
DR N-PSDB; ABK89163.
XX
PT Novel jAZF1, jjAZ1 or jAZF1/jjAZ1 polypeptides useful as immunogens or
PT antigens to raise or test anti-jAZF1, jjAZ1 or jAZF1/jjAZ1 antibodies
PT -
XX
PS Claim 7; Fig 3; 76pp; English.
XX
CC The present invention relates to a new jAZF1 (juxtaposed with another
CC zinc finger), jjAZ1 (joined with jAZF1) or jAZF1/jjAZ1 polypeptide.
CC The methods of the invention can be used to identify a compound which
CC controls proliferation of endometrial stroma, by expressing jjAZ in the
CC presence of the compound, and determining whether the compound affects
CC expression of jjAZ. jAZF1, jjAZ1 or jAZF1/jjAZ1 polypeptides are useful
CC as immunogens or antigens to raise or test anti-jAZF1, jjAZ1 or
CC jAZF1/jjAZ1 antibodies. The invention can be used as bait proteins in a
CC two hybrid assay or three hybrid assay to identify other proteins which
CC bind or interact with jAZF1/jjAZ1-binding proteins. jAZF1, jjAZ1 or
CC jAZF1/jjAZ1 molecules are useful for identifying the origin of tumour
CC and as tumour marker protein to verify that a stromal tumour is from
CC endometrium. The antibody is useful for promoting or decreasing
CC fertility or pregnancy, and also for treating endometrial stromal
CC tumours. The present amino acid sequence represents the human
CC jAZF1/jjAZ1 fusion protein of the invention.
XX
SQ Sequence 776 AA;

Query Match 83.2%; Score 3231; DB 23; Length 776;
Best Local Similarity 96.0%; Pred. No. 2.5e-250;
Matches 622; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

```
Qy         92 KPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESH 151
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 EPTQIYRFLRTRNLIAPIFLHRTLTYMSHRNSRTNIKRKTFKVDDMLSKVEKMKGEQESH 188

Qy        152 SLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKK 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 SLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTLEVLLVKVCHKKRKDVSCPIRQVPTGKK 248

Qy        212 QVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETN 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 QVPLIPDLNQTKPGNFPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRREFNGMINGETN 308

Qy        272 ENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKKR 331
```

```
PT   with altered vernalization response, flowering time, leaf size and/or
PT   shape or shade avoidance response for maximized reproductive success
XX
PS   Disclosure; Page 79; 105pp; English.
XX
CC   Isolated nucleic acid sequences obtained from the VRN2 locus of a
CC   plant encode polypeptides which are capable of affecting one or more
CC   vernalization responses such as, flowering time, leaf size and/or
CC   shape or the shade avoidance response of a plant into which the
CC   nucleic acid is introduced. Introducing such sequences into plants
CC   to alter these characteristics maximises the reproductive success of
CC   the plant. This protein is encoded an isolated human sequence which
CC   has homology to the Arabidopsis thaliana VRN2 gene over a short region
CC   near the N-terminus.
XX
SQ   Sequence   803 AA;

  Query Match           100.0%;  Score 3885;  DB 21;  Length 803;
  Best Local Similarity  100.0%;  Pred. No. 7.8e-303;
  Matches  739;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          65 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSSSA 124

Qy          61 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         125 AAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSH 184

Qy         121 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         185 RNSRTNIKRKTFKVDDMLSKVEKMKGEQESHSLSAHLQLTFTGFFHKNDKPSPNSENEQN 244

Qy         181 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         245 SVTLEVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSLAVSSNEFEPS 304

Qy         241 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         305 NSHMVKSYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMT 364

Qy         301 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         365 VFDKNRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLR 424

Qy         361 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         425 WTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPN 484

Qy         421 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         485 ENRQKLRIFYQFLYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCHSRFIFNYVY 544

Qy         481 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         545 HPKGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPITHILVCRPKRTKASM 604

Qy         541 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         605 SEFLESEDGEVEQQRTYSSGHNRLYFHSDTCLPLRPQEMEVDSEDEKDPEWLREKTITQI 664

Qy         601 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         665 EEFSDVNEGEKEVMKLWNLHVMKHGFIADNQMNHACMLFVENYGQKIIKKNLCRNFMLHL 724

Qy         661 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         725 VSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITEEQNGTANGFSEINSKEK 784

Qy         721 ALETDSVSGVSKQSKKQKL 739
              |||||||||||||||||||
Db         785 ALETDSVSGVSKQSKKQKL 803
```

```
RESULT 2
ABG30842
ID   ABG30842 standard; Protein; 739 AA.
XX
AC   ABG30842;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Human joined with jAZF1 (jjAZ1) protein.
XX
KW   Human; jAZF1; juxtaposed with another zinc finger; jjAZ1; jAZF1/jjAZ1;
KW   joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;
KW   antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
KW   chromosome 17.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          92..93
FT                   /note= "Breakpoint for production of fusion protein
FT                   jAZF1/jJAZ1"
FT   Misc-difference 448..471
FT                   /note= "Encoded by CTGCATTGCCCTTGGTGTACTCTGAACTGCCGCAA-
FT                   ACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT.
FT                   C2H2 zinc finger domain"
FT   Region          521..538
FT                   /note= "Bipartite nuclear localisation signal"
XX
PN   WO200193805-A2.
XX
PD   13-DEC-2001.
XX
PF   04-JUN-2001; 2001WO-US17936.
XX
PR   02-JUN-2000; 2000US-209093P.
XX
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX
PI   Koontz J,  Sklar J;
XX
DR   WPI; 2002-575047/61.
DR   N-PSDB; ABK89162.
XX
PT   Novel jAZF1, jjAZ1 or jAZF1/jjAZ1 polypeptides useful as immunogens or
PT   antigens to raise or test anti-jAZF1, jjAZ1 or jAZF1/jjAZ1 antibodies
PT   -
XX
PS   Claim 5; Fig 2; 76pp; English.
XX
CC   The present invention relates to a new jAZF1 (juxtaposed with another
CC   zinc finger), jjAZ1 (joined with jAZF1) or jAZF1/jjAZ1 polypeptide.
CC   The methods of the invention can be used to identify a compound which
CC   controls proliferation of endometrial stroma, by expressing jjAZ in the
CC   presence of the compound, and determining whether the compound affects
CC   expression of jjAZ. jAZF1, jjAZ1 or jAZF1/jjAZ1 polypeptides are useful
CC   as immunogens or antigens to raise or test anti-jAZF1, jjAZ1 or
CC   jAZF1/jjAZ1 antibodies. The invention can be used as bait proteins in a
CC   two hybrid assay or three hybrid assay to identify other proteins which
CC   bind or interact with jAZF1/jjAZ1-binding proteins. jAZF1, jjAZ1 or
CC   jAZF1/jjAZ1 molecules are useful for identifying the origin of tumour
CC   and as tumour marker protein to verify that a stromal tumour is from
CC   endometrium. The antibody is useful for promoting or decreasing
CC   fertility or pregnancy, and also for treating endometrial stromal
CC   tumours. The present amino acid sequence represents the human JjAZ1
CC   protein of the invention. This sequence is encoded by the human jjAZ1
CC   gene located on chromosome 17.
XX
SQ   Sequence   739 AA;

  Query Match           95.5%;  Score 3709;  DB 23;  Length 739;
  Best Local Similarity  96.8%;  Pred. No. 9.4e-289;
  Matches  715;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37 ; Search time 91.058 Seconds
(without alignments)
1288.178 Million cell updates/sec

Title: US-09-874-162A-5
Perfect score: 3885
Sequence: 1 MAPQKHGGGGGGGSGPSAGS.....KALETDSVSGVSKQSKKQKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3885	100.0	803	21	AAB00066	KIAA0160 polypepti
2	3709	95.5	739	23	ABG30842	Human joined with
3	3231	83.2	776	23	ABG30843	Human jAZF1/jjAZ1
4	2043	52.6	388	22	AAU15978	Human novel secret
5	2043	52.6	388	24	ABU55047	Human novel polype
6	1497	38.5	289	22	AAU15958	Human novel secret
7	1497	38.5	289	24	ABU55027	Human novel polype
8	1453	37.4	292	22	AAU16416	Human novel secret
9	1453	37.4	292	24	ABU55485	Human novel polype
10	1177	30.3	955	22	ABB71635	Drosophila melanog
11	919	23.7	175	22	AAU16402	Human novel secret
12	919	23.7	175	24	ABU55471	Human novel polype
13	243.5	6.3	604	21	AAY53933	An Os-MPC1 protein
14	236	6.1	611	21	AAY53932	A MPC1 protein hav
15	230	5.9	498	21	AAG54168	Arabidopsis thalia
16	227.5	5.9	367	21	AAG54169	Arabidopsis thalia
17	221	5.7	445	21	AAB00060	VRN2 polypeptide.
18	217	5.6	440	21	AAB00061	VRN2 polypeptide.
19	190	4.9	692	24	ABP70524	Histone deacetylas
20	190	4.9	813	21	AAB01674	FIS2 protein seque
21	162.5	4.2	1390	23	ABG96292	Human ovarian canc
22	161	4.1	342	22	ABB68951	Drosophila melanog
23	159.5	4.1	339	21	AAG54170	Arabidopsis thalia
24	159	4.1	3969	15	AAR52971	Product of the cDN
25	158	4.1	427	22	ABB65442	Drosophila melanog
26	157.5	4.1	1315	22	ABG09961	Novel human diagno
27	154.5	4.0	2311	23	ABG95034	Human translocatio
28	153	3.9	549	22	ABB65068	Drosophila melanog
29	153	3.9	549	22	ABB66483	Drosophila melanog
30	153	3.9	738	19	AAW56163	New DNA sequence i
31	147	3.8	831	16	AAR80168	pMISS1 MiSP spider
32	146.5	3.8	722	20	AAY43379	M. catarrhalis str
33	146.5	3.8	1235	16	AAR65158	Potassium ion chan
34	146.5	3.8	1235	22	AAM51809	S cerevisiae TRK1.
35	146.5	3.8	1639	19	AAW54145	P. falciparum synt
36	146.5	3.8	1639	23	AAE29345	Plasmodium falcipa
37	146	3.8	907	22	ABU53073	Intracellular traf
38	146	3.8	941	22	ABU53072	Intracellular traf
39	146	3.8	953	22	ABU53069	Intracellular traf
40	146	3.8	2230	24	ABU07445	Protein differenti
41	145	3.7	1570	22	ABB71314	Drosophila melanog
42	145	3.7	1570	22	ABB71316	Drosophila melanog
43	144.5	3.7	615	20	AAY40099	Spider silk protei
44	144.5	3.7	953	22	ABU53071	Intracellular traf
45	144.5	3.7	1654	6	AAP50777	Sequence of the P1

ALIGNMENTS

RESULT 1
AAB00066
ID AAB00066 standard; Protein; 803 AA.
XX
AC AAB00066;
XX
DT 16-NOV-2000 (first entry)
XX
DE KIAA0160 polypeptide.
XX
KW Vernalization gene; VRN2; plant characteristic; flowering time;
KW leaf size; leaf shape; shade avoidance response; reproduction;
KW breeding; pollination; cultivation; human.
XX
OS Homo sapiens.
XX
PN WO200044918-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-GB00248.
XX
PR 28-JAN-1999; 99GB-0001927.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Dean C, Gendall A;
XX
DR WPI; 2000-499333/44.
DR N-PSDB; AAA47759.
XX
PT Isolated vernalization gene VRN2 is used to produce transgenic plants